UNIVERSITY OF MIAMI
NOV 3 0 1965
LIBRARY

industrial mathematics

1953

UNIVERSITY OF MIAMI
NOV 3 0 1965
LIBRARY

industrial mathematics

1953

EDITORIAL STAFF

C. A. PIPER, EDITOR-IN-CHIEF, BENDIX AVIATION CORP.
R. V. CHURCHILL, UNIVERSITY OF MICHIGAN
K. W. FOLLEY, WAYNE UNIVERSITY
DOUGLAS HUGHSON, PACKARD MOTOR CAR COMPANY
A. W. JACOBSON, WAYNE UNIVERSITY
F. L. MONCHER, VICKERS, INC.
A. E. ROACH, GENERAL MOTORS CORP.
R. A. ROGGENBUCK, FORD MOTOR COMPANY
H. J. SCULLEN, CONSULTING ENGINEER
W. F. VOGEL, WAYNE UNIVERSITY

VOLUME 4

Copyright 1954 by

INDUSTRIAL MATHEMATICS SOCIETY
5135 CASS AVENUE, DETROIT 2, MICHIGAN

The Industrial Mathematics Society wishes to express its sincere appreciation to the Bendix Aviation Corporation for providing facilities for publication of Volume 4 of this Journal.

AN AXIOMATIC DERIVATION OF A GENERAL S-N EQUATION

Leonard G. Johnson
Research Laboratories Div., General Motors Corp.

ABSTRACT

A plot of the fatigue life of a specimen versus the applied stress is known as an S-N diagram. Due to the fact that otherwise identical specimens can have widely different fatigue lives under the same applied stress, it becomes necessary to consider an entire S-N region, embracing all levels of failure probability, and not merely a single S-N curve. What is needed, therefore, is a general S-N equation containing, among other parameters, the probability. By precisely defining such concepts as static strength, endurance limit, and minimum life, and making some well founded assumptions, it is possible to arrive at a general form of the S-N equation by means of dimensional analysis. Such a derivation is contained herein, as well as appended notes on some important consequences of the derived equation.

INTRODUCTION

The need for a general equation of S-N diagrams has been keenly felt in recent years. In the earliest stages of designing parts to be submitted to repeated stress, engineers simply constructed a single line on log-log paper to represent the relationship between average, or median, life and stress. This line was obtained by testing a few specimens at several stress levels, and then computing the average life at each stress. These averages were plotted vs. the stress, and in spite of the erratic arrangement of points on log-log paper, it was assumed that a straight line served the purposes of engineering well enough.

Until relatively recently, the fact that fatigue life at a fixed stress is a stochastic variable having a definite distribution had been accepted only by statisticians, while engineers still were thinking that if specimens were carefully prepared there should be no variation in endurance at a given stress. At present, however, engineers also realize that scatter in fatigue life at a fixed stress is a phenomenon to be expected, and consequently they do run many experiments in order to determine fatigue life distributions of various materials under particular types of loadings. Once these test results are obtained, they are usually analyzed by classical

methods, assuming that either life or its logarithm is a normally distributed variable at each stress level. Each stress level is individually analyzed by assuming the sample moments to be good estimates of population moments. This procedure often leads to inconsistent results in the relative amounts of scatter at various stress levels. Sometimes a high stress is shown to have more scatter than a low stress in such analyses. The life to stress relationship consequently becomes obscured, and most analysts work on the assumption that a linear relation exists between logN and logS in their particular problem, in spite of the inconsistent arrangement of points brought about by classical statistical analysis. While a linear relation between log N and log S with a fixed slope is a usable approximation in a limited range of stresses, we still need a general relation which is valid at all stress levels, and which will tell us how much scatter can be expected at any stress. Most experts have concluded that the line of best performance in an S-N diagram is less steep than the line of worst performance. Why? Because, as they say, it is only reasonable to have more scatter at lower stresses than at higher stresses, and hence the two lines must converge as we move upward on the stress axis. But, do they actually converge toward each other in a linear fashion, and come to a point of intersection? Furthermore, is there any relationship between the rate of convergence and the endurance limit? These and other questions cannot be answered until we first establish the fundamental relation between stress, life, and probability of failure.

The present work is an attempt to arrive at this relation in an axiomatic fashion, wherein only a few definitions and assumptions are used as a foundation.

DEFINITIONS AND ASSUMPTIONS

DEFINITION I, (q-life): The q-life of a stressed specimen is the number of cycles for which the probability of fatilure is q. (Denoted by $N_{q,s}$)

DEFINITION II, (Characteristic life): The characteristic life of a stressed specimen is the number of cycles for which the probability of failure is $1 - 1/e = 0.632$. (Denoted by θ_s)

DEFINITION III, (Minimum life): The minimum life of a stressed specimen is the minimum number of cycles which any such specimen will endure under the given stress. Some authors call this the "Sensitivity limit". (Denoted by $N_{o,s}$)

DEFINITION IV, (Static strength): The static strength of a specimen is that stress at which the specimen has a minimum life of zero cycles and a characteristic life of one cycle. (Denoted by U)

DEFINITION V, (Endurance limit): The endurance limit of a specimen is that critical value of stress at or below which the specimen will never fail under repeated loading, i.e., the stress under which the specimen has a characteristic life of infinity. (Denoted by L).

Using the above definitions, we now proceed to build a theory of S-N curves based on the following assumptions:

ASSUMPTION I: Any q-life is a weighted average of the minimum life and characteristic life of the form

$$N_{q,s} = \left[1 - \alpha(q)\right] \cdot N_{o,s} + \alpha(q) \cdot \theta_s ,$$

where $\alpha(q)$ is independent of the three stresses L, U, and S.

It follows from this that for $0 < q < 1-1/e$, we have $0 < \alpha(q) < 1$,
and, that for $1-1/e < q < 1$, we have $1 < \alpha(q) < \infty$.

We shall justify assumption I, as well as arrive at a formula for (q), in appendix I. This assumption essentially states that a life distribution function at a given stress is determined by exactly two parameters, i.e., $N_{o,s}$ and θ_s, $\alpha(q)$ being the same for all stress levels.

ASSUMPTION II: Any S-N equation must be dimensionally homogeneous. This assumption is commonly adopted as an axiom in dealing with relations between physical quantities, and we adopt it here as well.

THE BASIC VARIABLES OF THE GENERAL S-N EQUATION

Every fatigue specimen of a given material has a definite endurance limit, L, and a static strength, U. The magnitude of the applied stress shall be denoted by S. Therefore, there are three stress variables entering into the S-N equation, i.e., L, U, and S.

The S-N equation is an expression for the q-life at stress S, i.e ., a formula for $N_{q,s}$ in terms of the stresses L, U, and S, and other necessary variables. These other necessary variables are determined from assumption I, which states that $N_{q,s}$ is a function of the three variables $N_{o,s}$, $\alpha(q)$, and θ_s. Since the S-N equation already contains L, U, and S,

we are permitted to take only two of the three variables $N_{o,s}$, $\alpha(q)$, θ_s in order to complete it. If we take $N_{o,s}$ and θ_s as additional variables, the S-N equation becomes

$$N_{q,s} = F(L, U, S, N_{o,s}, \theta_s), \quad \text{which contradicts}$$

assumption I in that it implies that $\alpha(q)$ is a function of L, U, and S. We therefore take either of the other two possible pairs from $N_{o,s}$, $\alpha(q)$, and θ_s as our additional variables. Let us take the pair $N_{o,s}$ and $\alpha(q)$. Then the S-N equation becomes

$$N_{q,s} = F\left[L, U, S, N_{o,s}, \alpha(q)\right] \tag{1}$$

From assumption I:

$$N_{q,s} = G(N_{o,s}, \alpha(q), \theta_s) \tag{2}$$

Hence, (1) and (2) imply that θ_s is a function of L, U, and S, which is a reasonable conclusion. We therefore adopt the form (1) as a permissible S-N equation.

In assumption I let us put S = U. Then we have

$$N_{q,U} = \left[1 - \alpha(q)\right] \cdot N_{o,U} + \alpha(q) \cdot \theta_U \ .$$

According to definition IV, $N_{o,U} = 0$, and $\theta_U = 1$.

$$\therefore N_{q,U} = \alpha(q) \ . \tag{3}$$

Therefore, the S-N equation (1) may be written as follows:

$$N_{q,s} = F(L, U, S, N_{o,s}, N_{q,U}) \ . \tag{4}$$

THE CONSEQUENCES OF ASSUMPTION II

The general S-N equation is seen to consist of six variables related functionally as indicated in (4). Thus,

$$N_{q,s} = F(L, U, S, N_{o,s}, N_{q,U}) \ .$$

In terms of length, force, and life, the dimensional matrix of this set of variables is

	$N_{q,s}$	L	U	S	$N_{o,s}$	$N_{q,U}$
length	0	-2	-2	-2	0	0
force	0	1	1	1	0	0
life	1	0	0	0	1	1

The rank of this matrix is two. Therefore, there are 6 - 2 = 4 independent dimensionless parameters in this system. They may be taken to be

$$\left(\frac{N_{q,s}}{N_{q,U}}\right), \left(\frac{N_{o,s}}{N_{q,U}}\right), \left(\frac{U}{L}\right), \text{ and } \left(\frac{S}{L}\right) .$$

Consequently,

$$\left(\frac{N_{q,s}}{N_{q,U}}\right) = \phi\left[\left(\frac{N_{o,s}}{N_{q,U}}\right), \left(\frac{U}{L}\right), \left(\frac{S}{L}\right)\right] , \text{ where the function } \phi \text{ is}$$

to be determined.

There are two boundary conditions which the function ϕ must satisfy. They are

1st. : For $S = U$, $N_{q,s} = N_{q,U}$.

$$\therefore \phi\left[\left(\frac{N_{o,U}}{N_{q,U}}\right), \left(\frac{U}{L}\right), \left(\frac{U}{L}\right)\right] = 1 .$$

2nd. : For $S = L$, $N_{q,s} = \infty$.

$$\therefore \phi\left[\left(\frac{N_{o,L}}{N_{q,U}}\right), \left(\frac{U}{L}\right), 1\right] = \infty .$$

A particular solution, satisfying these two conditions, is

$$\left(\frac{N_{q,s}}{N_{q,U}}\right) = \phi\left[\left(\frac{N_{o,s}}{N_{q,U}}\right), \left(\frac{U}{L}\right), \left(\frac{S}{L}\right)\right] = \left(\frac{N_{o,s}}{N_{q,U}}\right) + \left(\frac{U/L - 1}{S/L - 1}\right)^m , \tag{5}$$

where $m > 0$. (5) may be put into the form

$$N_{q,s} = N_{o,s} + N_{q,U}\left(\frac{U - L}{S - L}\right)^m , \tag{6}$$

which may be taken as a general S-N equation, provided that it can be made to fit experimental data. At the present time, there does not exist a sufficiently wide collection of fatigue life data from independent sources to enable us to verify (6) very thoroughly. However, a critical examination of (6) indicates that it is very flexible, and can be made to fit very closely any fatigue situation which may come up for consideration.

APPENDIX I

Justification of assumption I: A. M. Freudenthal and E. J. Gumbel* have shown that at a given stress S the fatigue life of a specimen has an asymptotic cumulative distribution function of the form

$$q = 1 - \exp\left[-\left(\frac{N_{q,s} - N_{o,s}}{\theta_s - N_{o,s}}\right)^b\right] , \quad (b > 0), \tag{7}$$

which has come to be known as "The Weibull Function", due to the empirical analyses which W. Weibull, of Sweden, carried out on the strength and life of materials by using this function as a basis. Assumption I can be justified completely from (7) simply by solving it for $N_{q,s}$. We have

$$\exp\left[-\left(\frac{N_{q,s} - N_{o,s}}{\theta_s - N_{o,s}}\right)^b\right] = 1 - q .$$

*Freudenthal, A. M. & Gumbel, E. J.: On the Statistical Interpretation of Fatigue Tests, Proceedings of the Royal Society, A, Vol. 216, p. 309, (1953).

Hence,

$$\exp\left[\left(\frac{N_{q,s} - N_{o,s}}{\theta_s - N_{o,s}}\right)^b\right] = \frac{1}{1-q}$$

or,

$$\left(\frac{N_{q,s} - N_{o,s}}{\theta_s - N_{o,s}}\right)^b = \ln\frac{1}{1-q} \quad , \text{ or } \frac{N_{q,s} - N_{o,s}}{\theta_s - N_{o,s}} = \left(\ln\frac{1}{1-q}\right)^{1/b} .$$

Thus,

$$N_{q,s} - N_{o,s} = (\theta_s - N_{o,s})\left(\ln\frac{1}{1-q}\right)^{1/b} .$$

Transposing the $N_{o,s}$ term on the left to the right, we finally obtain

$$N_{q,s} = \left[1 - \left(\ln\frac{1}{1-q}\right)^{1/b}\right] \cdot N_{o,s} + \left(\ln\frac{1}{1-q}\right)^{1/b} \cdot \theta_s \quad ,$$

which is assumption I with $\alpha(q) = \left(\ln\frac{1}{1-q}\right)^{1/b}$.

Hence, due to (3), $N_{q,U} = \left(\ln\frac{1}{1-q}\right)^{1/b}$.

APPENDIX II

Some Consequences of (6)

(a) The case of zero endurance limit and zero minimum life.

In case $N_{o,s} = 0$ and $L = 0$, the S-N equation (6) reduces to

$$N_{q,s} = N_{q,U}\left(\frac{U}{S}\right)^m \quad ,$$

which means that a plot of log $N_{q,s}$ vs. log S will be linear and of the same slope for every probability level q. This is what happens in the case of ball bearing fatigue.

(b) A formula for minimum life in terms of stress and characteristic life.

Since (7) holds for all values of S, it follows that

$$\frac{N_{q,s} - N_{o,s}}{\theta_s - N_{o,s}} = \frac{N_{q,U} - N_{o,U}}{\theta_U - N_{o,U}} = \frac{N_{q,U} - 0}{1 - 0} = N_{q,U} \quad .$$

Hence, $N_{q,s} - N_{o,s} = N_{q,U}(\theta_s - N_{o,s})$.

From (6): $N_{q,s} - N_{o,s} = N_{q,U}\left(\frac{U - L}{S - L}\right)^m$.

$$\therefore N_{q,U}(\theta_s - N_{o,s}) = N_{q,U}\left(\frac{U - L}{S - L}\right)^m$$

or, $\theta_s - N_{o,s} = \left(\frac{U - L}{S - L}\right)^m$

Hence, $N_{o,s} = \theta_s - \left(\frac{U - L}{S - L}\right)^m$, from which it can be seen that the minimum life is zero if, and only if, $\theta_s = \left(\frac{U - L}{S - L}\right)^m$.

EXPANSIONS IN SERIES OF NON-ORTHOGONAL EIGENFUNCTIONS

Robert Davies
Research Laboratories Division, General Motors Corporation
Detroit, Michigan

ABSTRACT

Usually eigenfunctions are orthogonal, and the orthogonality is required in order to obtain, by the customary schemes, an expansion in a series of eigenfunctions. For example, the orthogonality is used to expand in a Fourier series or in a series of Bessel, Legendre, or Tschebyscheff functions. However, some physical problems lead to eigenfunctions which are not orthogonal, and other methods must be used to obtain the eigenfunction expansion. The paper discusses some of the techniques which are available.

INTRODUCTION

In many cases it is possible to reduce a second-order partial-differential-equation boundary-value problem to an ordinary differential equation system by separating the variables.[1-4] Usually a Sturm-Liouville system,[5]

$$\frac{d}{dx}\left[k(x)y'(x)\right] + \left[\lambda g(x) - \ell(x)\right] y(x) = 0,$$

$$a_{11}y(a) + a_{12}y'(a) + b_{11}y(b) + b_{12}y'(b) = 0,$$

$$a_{21}y(a) + a_{22}y'(a) + b_{21}y(b) + b_{22}y'(b) = 0,$$

is obtained. In the equation, $k(x)$, $g(x)$, $\ell(x)$, the a's, and the b's are known and λ is an arbitrary constant. The function $y(x) \equiv 0$ is a trivial solution for all values of λ. There will be non-trivial solutions, the <u>eigenfunctions</u>, for only certain values of λ, the <u>eigenvalues</u>.

If

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} k(b) = \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} k(a),$$

the system is called self-adjoint. For two eigenfunctions, $y_i(x)$ and $y_j(x)$,

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \begin{vmatrix} y_i(a) & y_j(a) \\ y_i'(a) & y_j'(a) \end{vmatrix} = \begin{vmatrix} a_{11}y_i(a) + a_{12}y_i'(a) & a_{11}y_j(a) + a_{12}y_j'(a) \\ a_{21}y_i(a) + a_{22}y_i'(a) & a_{21}y_j(a) + a_{22}y_j'(a) \end{vmatrix} =$$

$$= \begin{vmatrix} b_{11}y_i(b) + b_{12}y_i'(b) & b_{11}y_j(b) + b_{12}y_j'(b) \\ b_{21}y_i(b) + b_{22}y_i'(b) & b_{21}y_j(b) + b_{22}y_j'(b) \end{vmatrix} = \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \begin{vmatrix} y_i(b) & y_j(b) \\ y_i'(b) & y_j'(b) \end{vmatrix}.$$

Consequently, for self-adjoint systems,

$$\begin{vmatrix} y_i(a) & y_j(a) \\ y_i'(a) & y_j'(a) \end{vmatrix} k(a) = \begin{vmatrix} y_i(b) & y_j(b) \\ y_i'(b) & y_j'(b) \end{vmatrix} k(b).$$

If λ_i and λ_j are the eigenvalues corresponding to the two solutions $y_i(x)$ and $y_j(x)$, then, integrating by parts,

$$(\lambda_i - \lambda_j) \int_a^b g(x)y_i(x)y_j(x)dx = \int_a^b \left[\lambda_i g(x)y_i(x)\right] y_j(x)dx - \int_a^b y_i(x) \left[\lambda_j g(x)y_j(x)\right] dx =$$

$$= -\int_a^b y_j(x) \frac{d}{dx} \left[k(x)y_i'(x)\right] dx + \int_a^b y_j(x)\ell(x)y_i(x)dx +$$

$$+ \int_a^b y_i(x) \frac{d}{dx} \left[k(x)y_j'(x)\right] dx - \int_a^b y_i(x)\ell(x)y_j(x)dx =$$

$$= -\left[y_j(x)k(x)y_i'(x)\right]_a^b + \int_a^b y_j'(x)k(x)y_i'(x)dx +$$

$$+ \left[y_i(x)k(x)y_j'(x)\right]_a^b - \int_a^b y_i'(x)k(x)y_j'(x)dx =$$

$$= -\begin{vmatrix} y_i(a) & y_j(a) \\ y_i'(a) & y_j'(a) \end{vmatrix} k(a) + \begin{vmatrix} y_i(b) & y_j(b) \\ y_i'(b) & y_j'(b) \end{vmatrix} k(b) = 0.$$

In case $\lambda_i \neq \lambda_j$, the integral itself is zero. If there is only one eigenfunction for each eigenvalue, then in case $\lambda_i = \lambda_j$, the functions $y_i(x)$ and $y_j(x)$ are equal. The integral $\int_a^b g(x)y_i^2(x)dx$ is positive if $g(x)$ is positive

in the interval (a, b); then each eigenfunction $y_i(x)$ can be normalized by dividing by the square root of the integral. For normalized eigenfunctions,

$$\int_a^b g(x)y_i(x)y_j(x)dx = \delta_{ij}\,, \tag{1}$$

in which δ_{ij} equals 1 if i = j; otherwise δ_{ij} is 0.

Equation (1) establishes the orthogonality of the eigenfunctions. Representing the integral $\int_a^b g(x)y_i(x)y_j(x)dx$ as an inner product (y_i, y_j), the equation (1) becomes

$$(y_i, y_j) = \delta_{ij}\,.$$

If it is desired to have an eigenfunction series

$$f(x) = \sum_{i=0}^{\infty} f_i y_i(x) \tag{2}$$

that converges to an arbitrary function f(x), then one can find the coefficients f_i by assuming that the series converges and that term-wise integration is justified:

$$\int_a^b g(x)f(x)y_j(x)dx = \int_a^b g(x)y_j(x)\sum_{i=0}^{\infty} f_i y_i(x)dx = \sum_{i=0}^{\infty} f_i \int_a^b g(x)y_i(x)y_j(x)dx =$$

$$= \sum_{i=0}^{\infty} f_i \delta_{ij} = f_j.$$

Consequently, the coefficients in the expansion (2) should be

$$f_i = \int_a^b g(x)f(x)y_i(x)dx. \tag{3}$$

In inner-product notation,

$$f_i = (f, y_i).$$

It has been proved that an expansion in eigenfunctions of a Sturm-Liouville system of the type at hand converges to $f(x)$ on the interval $a<x<b$, if $f(x)$ is piecewise differentiable. [6]

As an example of an eigenfunction expansion, the Sturm-Liouville system

$$y''(x) + \lambda y(x) = 0,$$

$$y(0) = 0,$$

$$y(\pi) = 0$$

yields the eigenfunctions

$$y_n(x) = \sqrt{\frac{2}{\pi}} \sin nx, \qquad \lambda_n = n^2 .$$

For the functions,

$$\frac{2}{\pi}\int_0^{\pi} \sin nx \sin mx \, dx = \delta_{nm} .$$

The corresponding eigenfunction expansion is the Fourier sine series

$$f(x) \sim \sum_{n=0}^{\infty} f_n \sqrt{\frac{2}{\pi}} \sin nx, \qquad f_n = \sqrt{\frac{2}{\pi}}\int_0^{\pi} f(x) \sin nx \, dx .$$

Expansion of the function in a non-orthogonal series. From time to time a physical problem leads to a modified Sturm-Liouville system for which the orthogonality condition (1) does not hold. Instead there is a condition of the type

$$\int_a^b g(x)y_i(x)y_j(x)dx - Ay_i(a)y_j(a) + By_i(b)y_j(b) = \delta_{ij} . \tag{4}$$

It is assumed that the expression

$$\int_a^b g(x)y_i^2(x)dx - Ay_i^2(a) + By_i^2(b)$$

is positive for all of the eigenfunctions and that they can be normalized by dividing by its square root. As before, an inner product can be defined,

$$(y_i;y_j) = \int_a^b g(x)y_i(x)y_j(x)dx - Ay_i(a)y_j(a) + By_i(b)y_j(b),$$

$$(y_i;y_j) = \delta_{ij} .$$

From analogy with the previous case, the coefficients in an expansion

$$f(x) = \sum_{i=0}^{\infty} f_i y_i(x) \tag{5}$$

should be [7]

$$f_i = [f(x);y_i(x)] . \tag{6}$$

To derive the formula (6), assume that the series converges and that term-wise integration is justified.

$$\int_a^b g(x)f(x)y_j(x)dx - Af(a)y_j(a) + Bf(b)y_j(b) =$$

$$= \int_a^b g(x)y_j(x)\sum_{i=0}^{\infty} f_i y_i(x)dx - Ay_j(a)\sum_{i=0}^{\infty} f_i y_i(a) + By_j(b)\sum_{i=0}^{\infty} f_i y_i(b) =$$

$$= \sum_{i=0}^{\infty} f_i \left[\int_a^b g(x)y_i(x)y_j(x)dx - Ay_i(a)y_j(a) + By_i(b)y_j(b) \right] +$$

$$= \sum_{i=0}^{\infty} f_i \delta_{ij} = f_j .$$

Consequently the coefficients for the expansion (5) should be

$$f_i = \int_a^b g(x)f(x)y_i(x)dx - Af(a)y_i(a) + Bf(b)y_i(b) , \tag{7}$$

which is the equivalent of the equation (6). The convergence of series of non-orthogonal eigenfunctions has been established for only a few cases. [8-10]

Expansion of the derivative in an orthogonal series. Often a physical problem leads to non-orthogonal eigenfunctions because the boundary conditions contain second derivatives. Differentiating the differential equation once will yield a system with orthogonal eigenfunctions if the coefficients in the equation are constants and if the undifferentiated function does not occur in the boundary conditions. Consequently, the derivatives of the original non-orthogonal eigenfunctions are orthogonal. In other situations, a linear combination of the original eigenfunctions and their first derivatives, or a higher derivative, may be orthogonal.

An example of orthogonal derivatives is furnished by the longitudinal vibration of a rod which is clamped at one end, is attached to a mass at the other end, and is otherwise free from external forces. The differential equation is [11]

$$\frac{\partial^2 w}{\partial t^2} = \frac{E}{\rho} \frac{\partial^2 w}{\partial x^2} ,$$

in which $w(x, t)$ is the displacement of the rod at location x and time t, and E and ρ are constants for a particular rod. The boundary conditions are

$$w(0, t) = 0, \qquad M w_{tt}(\ell, t) + E w_x(\ell, t) = 0,$$

in which M is the attached mass and ℓ is the length of the rod. If the differential equation is satisfied at $x = \ell$, then the second boundary condition can be written

$$\frac{M}{\rho} w_{xx}(\ell, t) + w_x(\ell, t) = 0.$$

Upon separating the variables, the ordinary differential system becomes

$$y''(x) + \lambda y(x) = 0,$$

$$y(0) = 0,$$

$$A y''(\ell) + y'(\ell) = 0,$$

in which A is a constant. The system does not have orthogonal eigenfunctions, but on defining

$$z(x) = \frac{1}{\sqrt{\lambda}} y'(x),$$

$z(x)$ must be a solution of

$$z''(x) + \lambda z(x) = 0,$$

$$z'(0) = 0,$$

$$Az'(\ell) + z(\ell) = 0.$$

Since a Sturm-Liouville system is obtained, the eigenfunctions are orthogonal. Representing them by $z_i(x)$,

$$\int_0^{\ell} z_i(x) z_j(x) dx = \delta_{ij} ,$$

and the expansion associated with a function $f'(x)$ is, from the equation (3),

$$f'(x) \sim \sum_{i=0}^{\infty} f_i' z_i(x), \qquad f_i' = \int_0^{\ell} f'(x) z_i(x) dx.$$

If $f'(x)$ is the derivative of a function $f(x)$ whose expansion in a series of the eigenfunctions $y_i(x)$ is desired, and if term-wise integration is justified, then

$$f(x) = \int_0^x f'(x) dx = \sum_{i=0}^{\infty} f_i' \int_0^x z_i(x) dx = \sum_{i=0}^{\infty} \frac{f_i'}{\sqrt{\lambda_i}} \quad y_i(x).$$

The series converges to 0 at $x=0$ and therefore $f(0)=0$. The desired coef-coefficients are $\frac{f_i'}{\sqrt{\lambda_i}}$. Integration by parts gives them in terms of the functions $y_i(x)$:

$$f_i = \frac{f_i'}{\sqrt{\lambda_i}} = \frac{1}{\sqrt{\lambda_i}} \int_0^{\ell} f'(x) z_i(x) dx = \frac{1}{\sqrt{\lambda_i}} \left[f(x) z_i(x) \Big|_0^{\ell} - \int_0^{\ell} f(x) z_i'(x) dx \right] =$$

$$= \frac{1}{\lambda_i} \left[f(\ell) y_i'(\ell) - \int_0^{\ell} f(x) y_i''(x) dx \right] = \int_0^{\ell} f(x) y_i(x) dx + A f(\ell) y_i(\ell).$$

The equation (7) yields the same form. However, if the equation (7) is used, convergence of the series can be proved easily only for the few

cases treated in the literature. The method used here ensures convergence in case $f'(x)$ is piecewise differentiable and if the integration is justified. There are special techniques available for justifying the termwise integration of eigenfunction expansions. [9]

REFERENCES

1. Sokolnikoff, I.S., and Sokolnikoff, E. S., Higher Mathematics for Engineers and Physicists, McGraw-Hill Book Company, New York, 1934, p. 313.

2. Webster, A. G., Partial Differential Equations of Mathematical Physics, Hafner Publishing Company, New York, 1947, p. 102.

3. Madelung; E., Die Mathematischen Helfsmittel des Physikers, Dover Publications, New York, 1943, p. 192.

4. Tamarkin, J. D., and Feller, W., Partial Differential Equations, Brown University, Providence, R. I., 1941, p.65.

5. Ince, E. L., Ordinary Differential Equations, Dover Publications, New York, 1944, p. 217.

6. Birkhoff, G. D., Boundary Value and Expansion Problems of Ordinary Linear Differential Equations; Transactions of the American Mathematical Society, vol. 9 (1908), p. 390; Collected Mathematical Papers, American Mathematical Society, New York, 1950, p. 31.

7. Langer, R. E., Fourier's Series. The Genesis and Evolution of a Theory, The First Slaught Memorial Paper, Supplement to American Mathematical Monthly, vol. 54 (1947).

8. Langer, R. E., A Problem in Diffusion or in the Flow of Heat for a Solid in Contact with a Fluid, Tohoku Mathematical Journal, vol. 35 (1932), p. 260.

9. Churchill, R. V., Expansions in Series of Non-Orthogonal Functions, Bulletin of the American Mathematical Society, vol. 48 (1942), p. 143.

10. Bauer, W. F., Modified Sturm-Liouville Systems, Quarterly of Applied Mathematics, vol. 11 (1953), p. 273.

11. Love, A. E. H., A Treatise on the Mathematical Theory of Elasticity, Dover Publications, New York, 1944, p. 428.

ON SOME NON-LINEAR PARTIAL DIFFERENTIAL EQUATIONS*

S. D. Conte - Dept. of Mathematics, Wayne University

ABSTRACT

This note presents a method of obtaining exact solutions to certain types of quasi-linear partial differential equations. Given a partial differential equation $L(u) = 0$, a transformation $u = F(\theta)$ is introduced. The function θ is chosen as the solution of a certain linear partial differential equation, while the function F satisfies an ordinary non-linear equation. The proper choise of F and θ frequently leads to a solution of $L(u) = 0$. Several illustrations are given.

The usual linear theory used to describe phenomena in the fields of elasticity, hydrodynamics, aerodynamics, and heat conduction frequently fails to yield satisfactory results. In such cases recourse must be made to non-linear equations, but here the ordinary methods of obtaining solutions to linear equations, such as separation of variables and transform methods, are no longer effective. Indeed, exact solutions of non-linear equations are very rare. The purpose of this paper is to present a method of obtaining exact solutions to certain classes of quasi-linear partial differential equations.

If we are dealing with a partial differential equation whose dependent variable is $u(x, t)$, the method in its simplest form consists of introducing a transformation $u = F\left[\theta(x, t)\right]$ and choosing $\theta(x, t)$ as the solution of a certain linear equation. The function $F(\theta)$ will usually satisfy an ordinary non-linear equation with θ as the independent variable, the non-linearity in the original partial differential equation being transferred to the function F. The proper choice of F and θ will sometimes yield a solution of the equation in u. The method can best be explained by illustration.

Let us consider the equation

$$u_t = k\, u_{xx} + A(u, u_x) \tag{1}$$

---

* Read at the regular meeting of the Society, March 12, 1953.

where differentiation is indicated by subscripts. It is assumed that k is a constant and A is a non-linear function of u and u_x. This may be interpreted as a generalized equation of one-dimensional heat flow. Some important cases of (1) which have been investigated are those with $A = e^u$, $e^{-1/u}$, uu_x. If $A = uu_x$, for example, (1) is the equation of heat conduction in a moving medium, or again (1) with $A = uu_x$ may be interpreted as a model equation for turbulent flow in hydrodynamics as J. Burgers[1] has shown. We will consider in more detail equation (1) with $A = P(u)\, u_x^2$,

$$u_t = k\, u_{xx} + P(u)\, u_x^2 , \tag{2}$$

where P(u) is any function of u, linear or non-linear. We seek a solution of (2) in the form $u = F\left[\theta(x,t)\right]$, the functions F and θ being at this point unspecified. Then

$$u_x = F'\theta_x , \quad u_t = F'\theta_t ,$$

$$u_{xx} = F'\theta_{xx} + F''\theta_x^2 ,$$

and substitution in (2) yields

$$F'\theta_t = k(F'\theta_{xx} + F''\theta_x^2) + P(F)F'^2\theta_x^2 .$$

This may be rewritten

$$F'\left[\theta_t - k\theta_{xx}\right] = \theta_x^2\left[kF'' + P(F)F'^2\right] . \tag{3}$$

Now equation (3) will be formally satisfied if we take θ and F to satisfy the respective equations

$$\theta_t - k\,\theta_{xx} = 0 , \tag{4}$$

$$kF'' + P(F)F'^2 = 0 . \tag{5}$$

Of course, (4) is the linear equation of one dimensional heat flow, and its general solution is well known[2]. Furthermore, (5) is an ordinary non-linear equation whose solution can be obtained readily. Indeed, if we divide (5) by F' and k, we obtain

$$\frac{F''}{F'} + \frac{P(F)}{k}\, F' = 0 ,$$

an equation which is separable and whose solution is

$$c_1\theta + c_2 = g(F) = \int \exp\left[\frac{1}{k}\int P(F)dF\right]\, dF .$$

In the simplest case, P(u) equal to a constant, the equation (1) is of considerable importance and its solution is given by

$$u(x,t) = F\left[\theta(x,t)\right] = k \ln\left[c_1\theta + c_2\right],$$

as J. Cole[3] has shown. It may be worth while to consider one case in more detail. For this purpose, consider (2) with $P(u) = 1/u$. Equation (2) may be written

$$uu_t = k\, uu_{xx} + u_x^2 \quad , \tag{6}$$

and equation (5) becomes

$$k\, FF'' + F'^2 = 0$$

The solution of the latter equation is

$$F = \left[\frac{1+k}{k}\,(c_1\theta + c_2)\right]^{\frac{k}{k+1}} \quad , \tag{7}$$

the constants c_1, c_2 being arbitrary. Thus if $\theta(x,t)$ is a solution of the heat equation (4), then $u = F(\theta)$ with F given by (7) provides formally a solution of the quasi-linear equation (6). More explicitly, consider the boundary value problem consisting of (6) with k = 1 and the conditions

$$\begin{aligned} &u(x,0) = f(x) \quad , && 0<x<1 \quad , \\ &u_x(0,t) = u_x(1,t) = 0 \quad , && t>0 \quad , \end{aligned} \tag{8}$$

where f(x) is a suitably restricted function. Then $u = \left[2\theta(x,t)\right]^{\frac{1}{2}}$ is a solution of this boundary value problem, the corresponding boundary value problem for $\theta(x,t)$ being

$$\begin{aligned} &\theta_t = \theta_{xx} \quad , && 0<x<1, \quad t>0 \quad , \\ &\theta(x,0) = \frac{1}{2}\left[f(x)\right]^2 = g(x) \quad , && 0<x<1 \quad , \\ &\theta_x(0,t) = \theta_x(1,t) = 0 \quad , && t>0 \quad . \end{aligned} \tag{9}$$

The system (9) may be interpreted as the boundary value problem for the temperature within an infinite slab whose faces x = 0 and x = 1 are insulated and whose initial temperature is $\theta(x,0) = g(x)$. It is known that (9)

possesses a unique bounded solution[2] which may be written

$$\theta(x,t) = c_o + \sum_{n=1}^{\infty} c_n \exp(-n^2 \pi^2 t) \cos n\pi x \qquad (10)$$

where

$$c_o = \int_0^1 g(x)dx = \frac{1}{2}\int_0^1 f^2(x)dx \ ,$$

$$c_n = 2\int_0^1 g(x) \cos n\pi x \, dx = \int_0^1 f^2(x) \cos n\pi x \, dx.$$

Hence a solution of (6) is given by

$$u(x,t) = \sqrt{2}\left[c_o + \sum_{n=1}^{\infty} c_n \exp(-n^2\pi^2 t) \cos n\pi x\right]^{\frac{1}{2}} \qquad (11)$$

and it may be verified directly that the conditions (8) are satisfied. Since θ, θ_x, θ_t, are continuous over the domain $R(0 < x < 1, \ t > 0)$, then so are u, u_x, u_t, provided that $\theta(x,t)$ does not vanish over the domain R. The necessity of this latter restriction may be seen in differentiating

$$u = \left[2\theta(x,t)\right]^{\frac{1}{2}} \ .$$

The uniqueness of the solution $u(x,t)$ follows from that of $\theta(x,t)$. A more detailed discussion of the properties of the solution (11) will not be attempted here.

Another important family of quasi-linear equations may be written

$$\Psi_{xx} + \Psi_{yy} = f(\Psi, \Psi_x) \ , \qquad (12)$$

where $f(\Psi,\Psi_x)$ is a non-linear function. When $f = \mu(\Psi_x^2 + \Psi_y^2)$, for example, the function θ is found to satisfy the two-dimensional Laplacian equation. A solution of (12) is then given by $u = -(\ln \theta)/\mu$

The method of attack proposed here may be applied to partial differential equations of any order, but it is obviously limited to equations with certain types of non-linear terms. When it is not directly applicable, a preliminary transformation will often render the equation tractable to this method.

REFERENCES

1. Burgers, J., "A Mathematical Model Illustrating the Theory of Turbulence," Advances in Applied Mechanics, Academic Press (1948).

2. Churchill, R.V., "Fourier Series and Boundary Value Problems," McGraw-Hill (1941).

3. Cole, J., "A Quasi-linear Parabolic Equation in Aerodynamics," Quar. of Appl. Math., Vol. 9 (1951-52).

ELEMENTS OF ATOMIC ENERGY: ANALYSIS AND SYNTHESIS*

Urner Liddel - Bendix Aviation Corporation

ABSTRACT

This tutorial paper traces the recent development of physics from the discovery of radioactivity to the atomic bomb. A list of the known particles indicates that over 15 different ones have now been identified or suspected.

In discussing the problem of atomic energy, it might be fruitful to use an historical approach to the problem, which is in effect the application of an analytical technique to the problem. Let us consider the characteristics of matter. The first property which was discovered was undoubtedly that of mass. The second characteristic which was discovered was that of electrical charge, and the third was the property of magnetism. These three fundamental characteristics remained the sole basic properties of matter until the very recent developments in considering the nucleus.

Matter exhibited itself in three phases: solid, liquid, and gaseous; but most of the properties could be described by considering the solid phase of matter. Newton developed his laws on the conservation of mass, momentum, etc., in consideration of this phase. By these concepts matter was hard and impenetrable.

In examining the conductivity of gases, canal rays (positive ions) were discovered; and it further appeared that electrical conduction was accomplished by negatively charged particles which were called electrons. Thomson calculated the radius of these particles as 2×10^{-13} centimeters by assuming that their rest mass was entirely due to electromagnetic charge.

Discovery of radioactivity by Becquerel in 1896 was the first evidence of breakdown in the inviolability of matter. Here, of course, it was discovered that matter spontaneously decomposed in such a fashion that mass

---

* Read at the regular meeting of the Society, April 9, 1953.

was no longer conserved. Rutherford's classical experiments on the scattering of radioactive particles by thin gold foils showed that there was actually a considerable amount of free space separating the atoms of supposedly solid matter. This space in fact occupied a major portion of the volume of the solid.

Planck in 1900 discovered that light, originally considered to be continuous, must be discontinuous. Thus, we are confronted with a situation where two previous concepts of continuum had to be abandoned for a new concept of particulate nature.

In 1905 Einstein published a paper on his special theory of relativity. Using the Lorentz-Fitzgerald formula for the contraction of electrons in fast motion and Newton's second law of motion, he arrived at his now famous equation, $E = MC^2$. This derivation can be accomplished in the following manner:

$$dE = F \cdot dx, \text{ since work = force} \cdot \text{distance}, \tag{1}$$

also,
$$F = \frac{d(mv)}{dt} \text{ from Newton's second law ;} \tag{2}$$

therefore,

$$dE = \frac{d(mv)}{dt}\, dx \, . \tag{3}$$

Since $\frac{dx}{dt} = v$, (3) may be rewritten

$$dE = v\, d(mv)$$

or
$$dE = v^2\, dm + mv\, dv \, . \tag{4}$$

the Lorentz-Fitzgerald equation for explaining the apparent contraction of rapidly moving bodies states

$$m = \frac{m_o}{\sqrt{1 - v^2/c^2}} \, , \tag{5}$$

where m = apparent mass,

m_o = rest mass ,

c = velocity of light .

Equation (5) may be rearranged as follows:

$$m^2(c^2 - v^2) = m_o^2 c^2 \quad . \tag{6}$$

Differentiating we have

$$c^2 \, dm - v^2 \, dm - mv \, dv = 0 \quad . \tag{7}$$

Substituting (4) in (7) we have

$$dE = c^2 \, dm \quad , \tag{8}$$

or $$E = mc^2 \tag{9}$$

If light were particulate in nature, it should show the properties of momentum, and this was, indeed, proved. These experiments led to the necessity of the concept of DUALISM. This concept is with us today; and, more and more, we are forced to describe the manner in which we examine a physical system to have confidence in the apparent experimental results. That is to say, the system itself appears to be affected by the mere process of examination.

A logical consequence of this concept of dualism was that matter in motion should have associated with it a particular wave length. This wave length was calculated by DeBroglie in 1923 to be

$$\lambda = \frac{h}{mv}$$

where h = Planck's constant, 6.6×10^{-27} erg sec. This theoretical concept was beautifully proved by the diffraction experiments of Davisson and Germer with electrons.

This concept of wave motion of matter immediately led to the development by Schroedinger of what is now known as the wave mechanics. By this procedure, one can calculate the probability that one will find a particle of matter at a given place at a given time, $|\psi|^2$, from the general equation:

$$\nabla^2 \psi - \frac{8\pi^2 \mu}{h^2} V\psi + \frac{4\pi\mu i}{h} \frac{\partial \psi}{\partial t} = 0$$

where

$$\Upsilon = f(x, y, z) \exp(2\pi i \omega t)$$

The last of what might be called the historical concepts was developed

by Heisenberg in 1927 and continues by the title of the "Uncertainty Principle." By this concept, it is not possible to define precisely the momentum and the coordinates of a particle at the same instant. If one precisely defines the momentum, the uncertainty in the coordinates becomes equal to h.

The equation is

$$\Delta p \ \Delta q = h$$

where Δp and Δq are uncertainties in the momenta and coordinates, respectively.

Now we are able to see the profound difference between the macroscopic and microscopic properties of matter. The classical mechanics deals entirely with macroscopic properties and is governed entirely by the concept of continuous fields. The new mathematics dealing with atomics is called quantum mechanics, which describes the microscopic properties of matter and deals entirely with discontinuous fields. A corollary deduction from quantum mechanics is its inability to differentiate between observation and fact. We must use the equations of probability to define a system.

As you know, the structure of atoms is today described by a concept originally due to Bohr. He conceived of atoms as miniature astronomical systems where the heavy nuclei contain the major portion of the mass of the system, and these nuclei are surrounded by orbital electrons in a planetary manner. In astronomical systems, we find only one planet in a particular orbit. In atomic structure, it is conceived that two electrons occupy each orbit. The characteristics of the electrons surrounding an atomic nucleus are uniquely described by quantum numbers. According to Pauli's exclusion principle, no two electrons in any atomic system may have identical quantum numbers. Also, the chemical properties of all atoms are uniquely described by the quantum numbers of their electrons.

If we wish to consider a system as composed of particles, we now use what is known as the Fermi-Dirac statistics. Here, the ψ terms in the Schroedinger equation are antisymmetrical on interchange of the particles. If the ψ terms are symmetrical, the system is continuous and may be composed only of photons instead of particles, and one then uses what is known as the Bose-Einstein statistics.

Let us turn from the discussion of the electronic structure of atoms and consider now the nucleus, which contains nearly all the mass and thus all the energy of the atoms. Until the end of the 19th century, atoms were

conceived to be hard spheres. The Bohr model placed essentially all the mass in the nucleus, and it was considered in 1920 to be composed of protons and electrons. The electrons provided the electrostatic force required to hold the protons together. Later it was concluded on theoretical grounds that there could be no electrons in the nucleus, even though it was known from Beta decay that electrons were ejected from the nucleus in radioactive atoms. The absence of electrons in the nucleus led to the concept of a new elementary particle called the neutron, which was supposed to be a relatively stable particle composed of one proton and one electron. These particles were discovered in 1932 by Chadwick from the bombardment of lithium by Alpha particles. This concept left completely unsolved the nature of the forces which held the particles together at the nucleus. The true nature of these forces occupies the attention of nuclear physicists even at the present time.

The Japanese physicist Yukawa conceived the notion that there was perhaps an additional particle present in the nucleus which had not yet been discovered. He computed the mass of this new particle on the basis of the uncertainty principle and the DeBroglie wave length. The wave length is assumed to be of nuclear dimensions, that is, 2×10^{-13} centimeters. The mass of the particle may be calculated to be approximately 200 electron masses.

$$\lambda = \frac{h}{mc}$$

$$2 \times 10^{-13} = \frac{6.6 \times 10^{-27}}{6.3 \times m \times 3 \times 10^{10}}$$

$$m = 176 \times 10^{-27} \text{ grams}$$

$$\cong 200 \text{ electron masses } (m_e = 9 \times 10^{-28} \text{ gm})$$

Such a particle was discovered in the debris resulting from cosmic radiation by Street, Stevenson, Anderson, and Neddermeyer in 1938, and here the picture remained static until after the war.

All matter is today considered to be composed of elementary particles. These particles are, by definition, elements of matter which contain no structure. Most of them are to be found in various nuclei or result from the interaction of supposedly elementary particles with nuclei. In order to consider these, it is necessary to define certain new terms.

(a) The rest mass of a particle is the mass which the particle would have if its velocity were zero. This term is of importance in nuclear reactions since they occur at high relative velocities and the momentum changes may be greatly different than if considered solely on the basis of interaction at low velocity.

(b) The mean life of a particle is the length of time which the particle would live on the average if it had zero velocity. It is obvious that if the particle is traveling with velocities near that of light, the life time may be greatly extended. The equation for this is identical in form to the equation for relative mass.

(c) The cross-section of an elementary particle for a given reaction is defined as that area which would theoretically result in collision with the same probability as disruptive interaction actually occurs. The geometrical cross-section of atomic nuclei may be calculated on the basis of the following formula:

$$\sigma = 1.4 \times 10^{-13} A^{\frac{1}{3}} \text{ cm}$$

where A is the atomic weight. Experimental cross-sections almost never coincide with the geometrical cross-section, varying from .001 to 10^6 times the geometrical cross-section.

(d) The spin of electrons is evidenced, for example, in the different physical properties of ortho and para hydrogen. In the nucleus, the spins of the elementary particles are assumed to be additive or subtractive.

(e) The isotopic spin formalism is perhaps the most difficult to describe and understand. It arose from a consideration of the problem of the deuteron. This nucleus is composed of one proton and one neutron. Consideration of the electrical and magnetic forces between these two particles leads to forces insufficient to account for the stability of the deuteron nucleus. Thus, it was assumed that neutrons and protons were different states of the same nucleon.

It must be emphasized that no real importance can be assigned to this isotopic spin formalism. It is useful in considering the theoretical approach to particle interaction studies. One danger of placing too much emphasis on this concept may be illustrated as follows. In 1920 electrons were assumed to be in the nucleus to account for Beta decay. In 1950 neutrons are assumed to be in the nucleus, likewise to explain Beta decay. One wonders what future concept of elementary particles in the nucleus may be derived to account for this same phenomenon.

The number and characteristics of elementary particles are changing at such a rapid rate that one hesitates to tabulate them without emphasizing that this is purely provisional tabulation. Hardly a group meeting of physicists occurs without a report of a new particle. Tables 1 and 2 of elementary particles represented the best state of experimental and theoretical knowledge in the spring of 1953.

In the early days of study of nuclear physics, one was constrained to utilize particles from natural radioactivity. This type of experimentation has largely ceased. With the development of the high energy particle accelerators, subsequent to the invention of the cyclotron, we are now able to get particles in the laboratory with energies up to 2.5 billion electron volts. Except for the neutrons discovered by Chadwick, all other unstable elementary particles have been discovered in cosmic radiation. The low intensity of cosmic radiation and the difficulty of carrying out experiments in altitudes up to 100,000 feet closely limited the experimental information available from these particles. With the accelerator, we are now able to generate nearly all elementary particles in the laboratory, so that it is unlikely that we will have to depend in the future upon cosmic radiation for our knowledge of elementary particle interactions. However, since cosmic ray particles have energies up to 1 million times that which man can generate in an accelerator, this field of study will last for some years to come.

There are many ways of detecting elementary particles. All of these ways depend upon the ionizing properties of charged particles. The only method available for detecting uncharged particles is to observe the recoil of a charged particle ejected on collision with the uncharged particle. The four primary tools in particle detection are:

Photographic Plates
Scintillation Counters
Ionization Chambers
Cloud Chambers

The photographic plate was the original method of detecting charged particles. It was used by Becquerel in his discovery of radioactivity. The technique has markedly advanced in the last few years and today remains the best method of observing characteristics of elementary particles. The scintillation counter was developed by Rutherford and is used in many forms today. We have progressed far from the zinc sulfide crystals used by Rutherford and today may construct our detecting "crystals" in any form desired by incorporation of a suitable phosphor in a transparent plastic. Ionization chambers and the more specialized Geiger counters are generally used for detection purposes only. The original cloud chamber was

PARTICLE	CHARGE	REST MASS	SPIN	MEAN LIFETIME	DECAY SCHEME
NEUTRINO	0	0	$\frac{1}{2}$	∞	—
ELECTRON	−	1	$\frac{1}{2}$	∞	—
POSITRON	+	1	$\frac{1}{2}$	∞	—
(ANTI PROTON)	−	1836	$\frac{1}{2}$	∞	—
PROTON	+	1836	$\frac{1}{2}$	∞	—
(ANTI NEUTRON)	0	1838	$\frac{1}{2}$	(750)	$P^- + e + \nu$
NEUTRON	0	1838	$\frac{1}{2}$	750	$P^+ + e + \nu$
PHOTON	0	0	1	∞	—
GRAVITON	0	0	2	∞	—

TABLE 1

—MESONS—

PARTICLE	CHARGE	REST MASS	SPIN	LIFETIME	DECAY SCHEME
NEG. MUON	−	210	$\frac{1}{2}$	2×10^{-6}	$e + 2\nu$
POS. MUON	+	210	$\frac{1}{2}$	2×10^{-6}	$p + 2\nu$
NEG. PION	−	275	0	2.6×10^{-8}	$\mu^- + \nu$
POS. PION	+	275	0	2.6×10^{-8}	$\mu^+ + \nu$
NEUT. PION	0	265	0	10^{-14}	2γ
τ MESON	±	975	?	$\ll 10^{-10}$	$\pi^{\pm} + \pi^+ + \pi^-$
τ MESON	0	~ 950	?	$\ll 10^{-10}$	$\pi^0 + \pi^+ + \pi^-$
κ MESON	±	~ 1100	? $\frac{1}{2}$	$\ll 10^{-10}$	$\mu^{\pm} + 2\nu$ (?)
χ MESON	±	~ 1530	? 0	$\ll 10^{-10}$	$\tau^{0,\pm} + \pi^{\pm,0}$
χ MESON	0	~ 1370	?	10^{-10}	$\tau^{\pm} + \pi^{\pm}$
NEG. V	−	~ 2300	?	10^{-10}	$N + \pi^- + \pi^0$?
POS. V	+	~ 2300	?	10^{-10}	$N + \pi^+ + \pi^0$?
NEUT. V	0	~ 2300	? $\frac{1}{2}$	3×10^{-10}	$N + \pi^+ + \pi^-$?
					$p + \pi^- + \pi^0$? (ν)

TABLE 2

developed in 1923 by Wilson and has progressed today to the large continuously recording diffusion cloud chambers.

In order to show some dramatic examples of interaction of elementary particles with matter, three cloud chamber photographs are reproduced herein. The first, Figure 1, shows the traversal of four lead plates 1/2 inch thick by a single elementary particle before it is slowed down to such an extent that it can interact with the lead. The large number of ionization particle tracks shown in the four lower sections illustrate the large number of consequences which can occur from a single primary interaction. This is perhaps more dramatically illustrated in the second picture, Figure 2, which shows an enormous electron shower generated in the wall of the cloud chamber. The last picture, Figure 3, illustrates a true nuclear explosion. Here, a single cosmic ray particle enters the photographic emulsion, seen in the upper right-hand corner of the picture. It collides with a single atomic nucleus, probably silver or bromine, and some 80 separate particles are ejected in this one explosion. It has been estimated that 10 ergs of energy were released in this one interaction.

Many nuclear physicists agree that the greatest need in the nuclear field at the present time is a concept of the nucleus similar to that of Bohr for the atom. There are three models under discussion at the present time with differing degrees of support. The liquid drop model assumes that the particles are held together by nuclear forces with free motion inhibited by viscous flow of the particles. Thus, fission is said to occur when the surface tension of the liquid is insufficient to hold the elementary particles together any longer.

The powder model assumes that the particles may act essentially independently of each other and are held together essentially by gravitational type forces.

The shell model has received perhaps the widest acceptance at the present time. Here, it is assumed that the elementary particles are in energy states comparable to the electron shells in the outer orbits of the atom. This model well accounts for many energy changes found in nuclear interactions. In addition, it has been discovered that mesons, on approaching an atomic nucleus, apparently fall into an energy level on capture by the nucleus, without falling directly into the nucleus itself. Subsequent to this capture, the meson then goes into the nucleus with emission of characteristic radiation.

Philosophically, we are in a dilemma. By the uncertainty principle we are not allowed to discuss lengths less than 10^{-13} centimeters; yet the

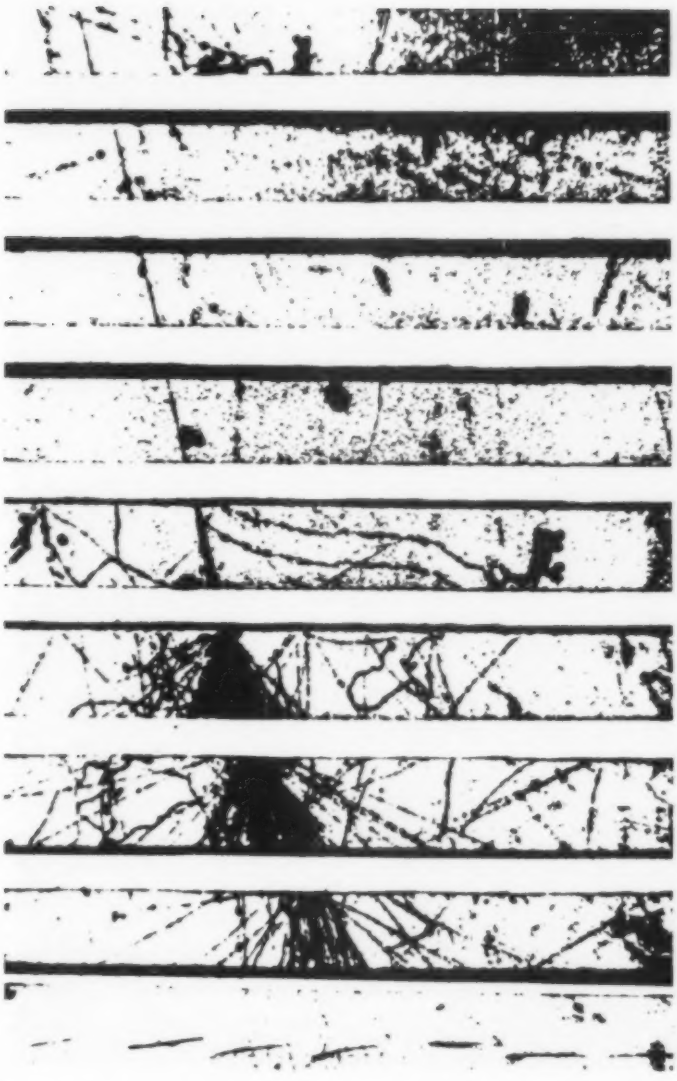


Figure 1 - High Speed Elementary Particles
Interacting with Lead Plates

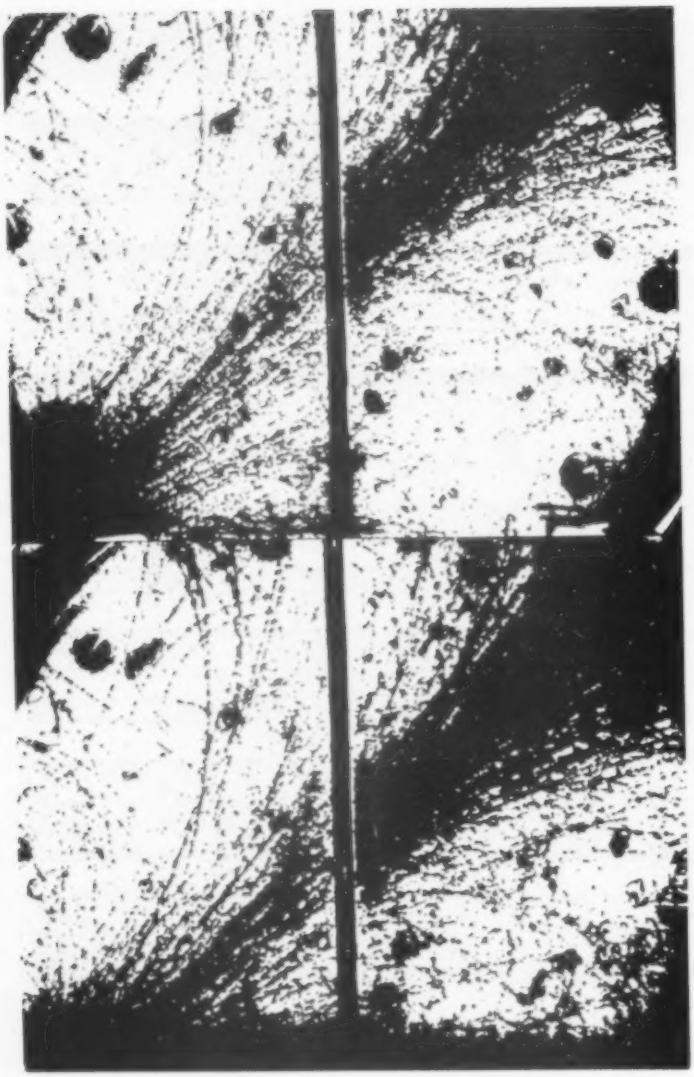


Figure 2 - Electron Shower in Cloud Chamber

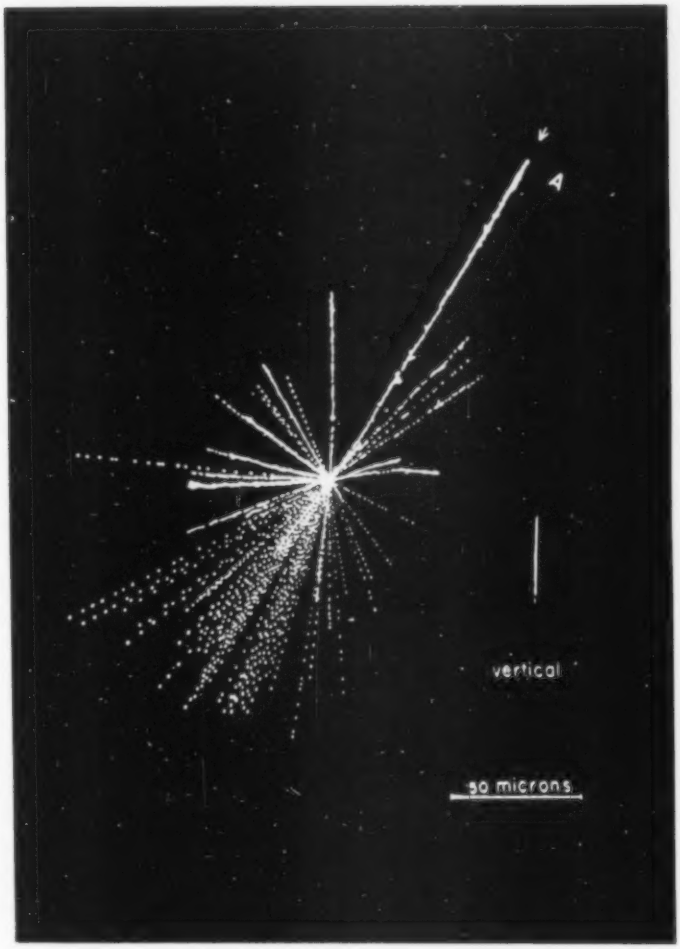


Figure 3 - Nuclear Explosion

dimensions involved in interaction of particles in the nucleus must occur at dimensions considerably less than this. It appears that we shall have to abandon the concept of particles completely within these small dimensions and say that they have reality only under isolated conditions.

One reads much in the newspapers concerning the utilization of nuclear energy. The energy involved in these discussions is that of the fission of Uranium 235. It can be shown that if one could obtain the complete fission of one gram of U-235, one would obtain an energy equivalent to 2.3×10^4 kilowatt hours. This is approximately equal to one million watts for one day. It is, of course, extremely difficult to obtain the complete fission of a single gram of material. One must, in fact, have many kilograms in order to obtain this quantity of energy. It should also be pointed out that this energy is essentially thermal energy and is not directly convertible into electrical energy. On the other hand, it may be interesting to point out that one obtains the order of .01 kilowatt hours per gram of coal, so that uranium provides nearly one million times as much energy as coal.

For the last 30 years, physicists have dreamed of the hydrogen condensation reaction in which four hydrogen atoms coalesce into one helium atom. This is believed by the astronomers to be the primary source of energy of the stars. If this reaction could be made to occur, one could obtain 1.25×10^5 kilowatt hours per gram of hydrogen. This condensation should have the possibility of being operated as a controlled reaction, free from the dangers of explosion and radiation hazards developed in uranium reactors. It is perhaps more than a fond hope that we may see this reaction as a source of energy within our lifetime.

In order to finish this cycle, it is perhaps well to note that if we could completely disintegrate matter into energy we would obtain from one gram of any kind of matter 2.5×10^7 kilowatt hours. It is obvious from these figures that any of the reactions which we are able to control utilize only a small fraction of the total energy available. Whether we are ever able to utilize total energy available from mass remains to be seen. The mesons previously discussed may hold a key not only to the forces holding the nucleus together but to a possible mechanism for the complete decomposition of matter into energy. Certainly, if we are ever able to utilize either of the last two methods of producing energy, we will no longer be dependent on fossil fuels for heat sources.

In summary then, I hope that it has been shown that matter is rather complicated -- the more complicated by our own inability precisely to measure the physical quantities involved. At the moment, the ephemeral particles called mesons have only significance for the theoretical and experimental physicists as items of great interest. However, one cannot

resist a strong belief that these particles will be effective in the power sources of the future.

THE INFLUENCE OF AUTOMATIC COMPUTING MACHINES ON MATHEMATICAL RESEARCH*

H. D. Huskey - Mathematics Department, Wayne University

ABSTRACT

The paper discusses the design and use of high-speed computers, and their influence on pure and applied mathematics. New computer design features aimed at more efficient use are discussed; an indication of future trends is given. Examples in number theory, matrix problems, linear programming, and Monte Carlo techniques are mentioned.

INTRODUCTION

The influence of automatic computers on mathematical research contains three aspects: (1) the opening of a new field in mathematics, namely the design and use of high-speed computers, (2) the influence of computers on trends in pure mathematics, and (3) the influence of computers in popularizing applied mathematics.

Present and prospective computing machines, have revitalized older studies and opened new vistas in many branches of mathematics. A field centering around the efficient use of existing computers and the design of improved machines has come into being. The development of machinery which will precisely carry out complex and intricate logical processes of millions upon millions of steps has led to the hailing of the era of the "mechanical brains" by one school, and to the vehement assertion of human "corner" on the ability to think by another. In point of fact, it is exceedingly difficult to define the word "think" in such a manner that an automatic computing machine can not qualify. However, all users of such equipment know that considerable thinking is necessary by the persons concerned before these machines can be properly utilized.

The advent of high-speed automatic computers has presented mathematicians with a challenge to use these machines as efficiently as possible.

---

* Read at the regular meeting of the Society, Nov. 14, 1952.

The use of computing machines may be considered from two viewpoints: programming techniques and computer facilities. In any general purpose digital computer it is possible so to program that the machine will behave like any other specified digital computer. In fact, it has been shown that a device with an infinite memory but only a rudimentary arithmetic unit can be caused to imitate any other computer. This means that the significant question is not whether a given computer can do a given problem but merely whether the computer is efficient on that problem. This efficiency depends partly upon what special facilities have been built into the computer and partly upon what programming techniques have been developed by the people who use the computer. Thus, this question of efficiency has to do both with programming techniques and with actual facilities built into the computing device.

COMPUTER DESIGN

First consider the computer facilities of the machines now in operation. In the initial machines the emphasis was on minimizing tubes and simplifying design in order to facilitate construction, check out, and installation. This led to computers having a minimum of commands, or at least a minimum variety of commands. One example of this is the National Bureau of Standards Western Automatic Computer, SWAC, in which there are only eight basic commands.

Another direction of development has been to let each digit in the command represent a particular elemental function. A combination of these functions then represents an arithmetic operation. Other computers have incorporated the arithmetic operations in the memory switching circuits. In these cases, a primary motivation was simplicity in design. There are computers now in operation using this principle, as for example, the ACE in England.

With the actual construction of automatic computers and their use in computational problems, there have appeared recently machines of more complicated design than these first models. This is partly because a major bottleneck in preparing problems for solution on high-speed computers is that of programming. Programming requires trained personnel and even then is a lengthy process, partly because of inherent complexity and partly because of the difficulty of detecting and removing errors. In the future it is hoped that machines will be designed which will permit simpler programming. One method would be to provide facilities for accomplishing substantially more erudite commands than are designed in the present line of computers. For example, vector operations could be ac-

complished with single commands. One machine has been recently designed to be especially useful on matrix problems.

At this point it may be of interest to note a few particular instructions, originated to facilitate coding. The repeat command in the ERA-1101 computer is designed to expedite obeying a group of commands a specified number of times.

The "B" register used in the Ferranti computer designed at Manchester University, England, allows a given set of commands to be applied to various sets of numbers without modification of the commands. Another aid to coding is the "relative address" system in the National Physical Laboratory's ACE (pilot model). Here parts of the memory can be precessed relative to other parts, so that a given set of commands can be applied to sequences of numbers with only the location of the commands being changed.

The record command of the UNIVAC and the subroutine transfer commands of the CIRCLE computer facilitate the entry into subroutines from the main problem routine. The problem is: how does the computer return to the proper place in the main routine after doing a subprogram operation such as a sine or cosine? The record command, or subroutine transfer specifies this in one word, whereas in other computers two words may be required.

A more recently designed computer, the DYSEAC at the National Bureau of Standards, incorporates some interesting commands. It has a summation command which sums a sequence of numbers in the memory. To facilitate floating point computations a "justify" or normalizing command shifts the number into standard form. (In a floating point operation the digits of one word position in the computer are divided into two groups. One group represents the mantissa or significant figures of the number, and the other the exponent of the base. For example, in a decimal computer +123456789 might represent the number $+0.1234567 \times 10^{39}$, the last two digits representing the exponent plus 50.) There are two command counters either of which may be used by any command. This facilitates subroutine entry and return to the main routine. A flexible system of storing and resetting these command counter contents is supplied by a so-called "file" command. A breakpoint file command provides these same facilities but with optional use depending upon the manual setting of a switch.

PROGRAMMING

Let us next turn our attention to programming techniques. Here is a subject that lies in-between the engineering and design of computers, and

the use of them by mathematicians. It seems preferable to classify programming as a mathematical subject since it is a necessary adjunct of operating a computing system. The operation of a computing system is primarily a mathematical activity, whether it be involved in engineering design or in number theory. The history and development of programming techniques is comparable to that given above concerning extra facilities in computing machines.

A high-speed automatic computer is such a complex device that methods of using a given computer continue to improve and develop even without any change in the logical or electronic aspect of that computer. For example, just recently there have been devised so-called post-mortem checking routines which process other routines in order to find out what is wrong with them. Work is also being done on floating or relative address systems that simplify the bookkeeping processes involved in laying out a program of computation. In these floating address systems the actual instructions involved in carrying out the computation are classified into various groups, and thus, in laying out the program it is only essential to make sure that the commands in each group are laid out in the proper order. Commands can be added to each group without upsetting the general arrangements of the other groups. Programming techniques of this sort have simplified greatly the preparation of problems for computation on high-speed machinery.

In this direction mathematicians may look forward to computing machines for which a problem can be essentially programmed by typing the formula that is to be evaluated. There would be no worries about where information is to be stored, or what logical patterns are to be involved in the computation. This sounds quite wonderful, of course, but on the other hand, it is doubtful if any of the purely mathematical preparation that must precede the running of a problem can be by-passed. For example, someone must decide what mathematical method is to be used, and how many significant digits are to be carried. Once these questions are decided, it is a simple matter to write down equations representing the processes involved. The equations can then be typed into the computer. Next the computer must interpret these equations and produce a list of commands, which when processed by the computer will give the answers to the desired problem.

Now consider the effect of automatic computer programming on developments in logic. One of the questions, of course, is how does one write down the logical equations that represent a computer design? This operation may involve multiple valued logical functions, and more particularly, the equations of a computer involve time as an independent variable. The techniques for writing the equations that represent the response

of a logical circuit are quite involved, and it is rather difficult to write such equations and make sure that they are correct. In general, most of the design of digital computing machines has up to this time at least been done by a trial and error, or brute force method. But perhaps techniques will be developed in applying logics to this field so as to derive a straightforward procedure for arriving at the optimum design of an electronic computer or similar device.

APPLICATIONS TO PURE MATHEMATICS

Consider now the field of pure mathematics. One of the first types of problems persons checking out computers like to try is factoring problems. This is because here one can specify quite large scale problems requiring a long running time, usually knowing the answer in advance. For example, assume the computing machine is instructed to divide 2^{18} by 2. As it is known that the answer is 2^{17}, the computer can be instructed to start subtracting 2's from 2^{18} and to tally how many times it does this. It will run on for quite a long period of time. At the end of the process, by looking at the remainder which should be 0, one can check that the computer probably ran correctly for a large number of operations.

Worthwhile applications of computing devices have been made in the field of factoring large primes. Of particular interest is the investigation of Mersenne primes that has been carried on as an incidental project at the Institute for Numerical Analysis in Los Angeles. (One interesting point about this problem is that it represents one case where a non-professional coder prepared a problem, and the code turned out to be correct in every detail.) The purpose is to study numbers of the form 2^p-1, and to find out for what values of p these numbers are prime (Mersenne primes). In general, the program is set up to handle numbers for p's as large as 2293. The limitation on the size of the number involved is the amount of room in the high-speed memory to store the number. Also for numbers much larger than this, computation time becomes a limiting factor. But to summarize in brief the results that were obtained, the table of Mersenne primes was extended by the addition of five new primes, including the largest known prime numbers at the present time.

The interest in Mersenne primes arises from their relation to perfect numbers, a perfect number being one which is equal to the sum of all its exact divisors; e.g. $28 = 1+2+4+7+14$. The only known general expression producing perfect numbers is due to one of Euclid's theorems which states that $2^p-1(2^p-1)$ is perfect if 2^p-1 is prime. Mersenne alleged in 1644 that 2^p-1 is prime for certain numbers, p, namely, 2, 3, 5, 7, 13, 17, 19, 31, 67, 127 and 257. Subsequently 67 and 257 were found not to result in

primes, and also it was discovered that Mersenne had omitted three numbers: 61, 89, and 107. In general, only in the last ten years have some of the final numbers been laboriously checked. For example, Dr. D. H. Lehmer, spending some 700 hours on a desk machine, was able to show that $2^{257}-1$ was a composite number. The SWAC was able to reach the same conclusion in 48 seconds. The SWAC on its first try on this problem, picked up two new Mersenne primes, namely, $2^{521}-1$ and $2^{607}-1$. Subsequently, three more numbers were added, the largest being $2^{2281}-1$.

This investigation of large Mersenne primes has other qualities of interest to those preparing problems for high-speed automatic computing machines. First, note that the largest prime number investigated in this system has 2281 digits in its binary representation, and thus, requires considerable space to write down. It may be asked, "How can a computing machine with a limited number of registers for storing information be used to compute with respect to a number of this size, which requires most of the memory just to write down the number?" The answer illustrates the power of, and the need for, mathematical methods along with computing machine methods. In this particular case a theorem of Lucas is used to prove the primality of the numbers involved. Lucas investigated Mersenne primes, and it is quite likely that he became tired of dividing in order to find out if a particular number was, or was not, prime. He eliminated a great deal of the tediousness when he evolved the following theorem: To investigate the primality of a number such as 2^5-1, which is equal to 31, construct a Lucas series in which the first term is 4, and any other term is the preceding term squared minus 2, etc. For example, if the first term is 4, the second term is 16-2, or 14, the third term is 196-2, or 194. Actually, one no longer has to deal with 194, since one is only interested in this number relative to 31. Hence, it can be reduced, modulo 31, which means that instead of dealing with 194 one only has to deal with the remainder when divided by 31, namely 8. The 4th Lucas number is 8 squared -2, or 62, etc. Now, if 2^P-1 is prime the p-1 term of the Lucas series is divisible by 2^P-1. In this particular example, the fourth term is equal to 62 which is divisible by 31; therefore, 31 is a prime number. This theorem was used in checking out the primality of the larger primes, and the important thing is that the machine had to store only the reduced number modulo 2^P-1. This example illustrates the power of advanced theoretical developments used in conjunction with improved computing methods and equipment.

High-speed computing techniques have been applied to other problems, for example, the Riemann Zeta hypothesis. No one knows whether this hypothesis is true or not - no one has been able to prove it - no one has been able to find a counter example. Many numbers have been investigated, so if counter examples exist, they are known to involve rather large

numbers. Certainly one of the possibilities of automatic computing machines is their use to discover counter examples to such theorems. However, to the author's knowledge, no one has yet used a computing machine actually to prove a mathematical theorem of this sort, although it has been proposed that possible propositions, depending upon a reasonably simple set of axioms, could be proved by the computer simply by exhausting all the possibilities of the true and false statements.

COMPUTERS IN APPLIED MATHEMATICS

The field of applied mathematics, as considered in relation to computing machines, may be divided into two different categories, statistical and engineering.

In the statistical field the interest of the Bureau of the Census in the tabulation of census data has perhaps spurred the development of computing machines as much as any other activity, in that this Bureau was one of the original government agencies to encourage private industry to build a computer for this particular purpose. The result of their interest was a study contract with the Eckert-Mauchly Computer Corporation which ultimately resulted in the production of a UNIVAC to be used by the Census Bureau. This machine has been in operation for several years now and has been used to tabulate census data on a reasonably large scale. In fact, the 1950 census for a number of the most populated states was processed on it. In general, it takes approximately 9 months for the Census Computer to process the census data on 15 million people.

Tabulation of such data is quite similar to the problems that arise in industry and business applications with respect to the amount of actual computing that must be done. In general, these problems involve tabulations or sorting of data in accumulating totals of items in various categories. Thus, machines designed for this purpose need high-speed input-output equipment, such as magnetic tape units, that will supply information to the computer at a high rate of speed.

The first high-speed electronic computer, the ENIAC, was built at the University of Pennsylvania with the support of the Army Ordnance Department. Their interest was primarily a good method of producing firing tables and the ENIAC was designed particularly for the solution of this problem. In fact, it can compute the trajectory of a shell in less time than the shell spends in flight. With the advent of other high-speed electronic computers there has been more interest in trajectory problems from the point of view of using higher order equations to represent the motion of the objects.

One of the problems that has been of interest to the aircraft industry is the vibration of airplanes. In general, the study of such vibrations involves studying flutter matrices to find their eigenvalues. A great deal of work has been done at the Institute for Numerical Analysis to find better methods of solving systems of linear equations, of finding eigenvalues for matrices, or for inverting matrices. These methods range from study of simple elimination methods to rather elegent methods involving steepest descent techniques. Steepest descent methods involve approximating a vector solution to the system and modifying this vector in a direction which minimizes a residual, or the sum of the squares of the residuals, when this vector is substituted in the system of equations.

A new branch in applied mathematics that has developed hand in hand with the development of high-speed automatic computation is that of linear programming. The ideas and concepts of linear programming are not new, but in general the problems that are involved in this field have so many equations, or inequalities, and so many unknowns, that it was impractical to work out examples before the advent of high-speed electronic computation.

With the advent of high speed computing, so-called Monte Carlo methods became almost a fad among mathematicians. There is a correspondence between certain sequences of random events and certain types of differential equations. Thus, the solution to certain problems involving sequences of random events can be found by solving the corresponding differential equation. Conversely, certain differential equations can be solved by computing sequences of random numbers and processing them in a prescribed way.

This can be illustrated by a problem called the Gambler's Ruin. Two people G and H play a game involving turns. The probability that G wins on any turn is a constant p, and the probability that H wins is q. The stake at each turn is h. What is the probability that G ruins H if their combined fortune is b?

Let x(t) be the fortune of G at any time and let $g_m(x)$ be the probability that G ruins H in not more than m turns.

Consider two cases: (1) G wins next turn and ruins H in not more than m-1 turns, (2) G loses next turn and ruins H in not more than m-1 turns. Therefore

$$g_m(x) = pg_{m-1}(x+h) + qg_{m-1}(x-h),\ 0 < x < b,\ g_m(0) = 0,\ g_m(b) = 1.$$

By letting $m \to \infty$ and $h \to 0$ a differential equation of the form

$$\frac{d^2g}{dx^2} + 2\alpha\frac{dg}{dx} = 0$$

where $\lim_{h \to 0} \frac{p-q}{qh} = 2\alpha$, is obtained.

The fraction of the times that G ruins H in these games may give a solution to the differential equation.

Numbers with sufficiently random properties may be generated in a high-speed computer, and in effect, cause the computer to play a number of games such as described above.

Similar "random walk" techniques can solve La Place's equation at a given point in a region with prescribed boundary values. Most other methods of solving such equations give answers over the whole region, and may involve much more computing than that needed if the value of the solution at one point is all that is desired.

In closing, I find the following points are worth emphasizing. First there are many extremely interesting problems which by presently known methods would require years of computing even on the fastest electronic computing equipment available. For these problems new methods are needed. The second point is that the process of automatic computation is almost an "art" at the present time. The time used to prepare and check problems competes with the time required to run them. Improvements both in computers and in techniques of using these computers are required.

This leads to a final point that although numerous commercial companies are delivering automatic computing equipment to all types of customers all over the country there is very little provision for training people to use these machines efficiently. However, Wayne University is pioneering in the development of a comprehensive training program, and it is the author's hope that in the next few years there will be a substantial increase in this type of training activity.

BIBLIOGRAPHY

Curtiss, John H., "Sampling Methods Applied to Differential and Difference Equations, "Reprint of a paper presented at the Seminar on Scientific Computation, International Business Machines Corporation, November 1949.

Huskey, H. D., "Charactristics of the Institute for Numerical Analysis Computer, "Mathematical Tables and Other Aids to Computation, IV, No. 30, pp. 103-108; April 1950.

Huskey, H. D., "Electronic Digital Computing in England, "Mathematical Tables and Other Aids to Computation, III, No. 23, pp. 213-216; July 1948.

Leiner, Alan L., "System Specifications for the DYSEAC," National Bureau of Standards Report 1951, 30 pp.; September 1952.

Reid, Constance, "Perfect Numbers, "Scientific American, Vol. 188, No. 3, pp. 84-86; March 1953.

Turing, A. M., "On Computable Numbers, With an Application to the Entscheidungsproblem, "London Math. Soc. Proc., s. 2, Vol. 42, pp. 230-265 (1937).

LINEAR ANALYSIS OF FEEDBACK-CONTROL SYSTEMS*

E. Calvin Johnson
Bendix Aviation Corporation, Research Laboratories

ABSTRACT

This tutorial paper presents the mathematical background underlying the analysis of feedback systems operating in the linear region. Classical and transform methods of solving the equations are taken up and some discussion devoted to steady state frequency response and its relation to system stability.

1. NONLINEARITY

As the subject indicates, this discussion is confined strictly to linear systems. Actually, all physical systems, if pushed hard enough or examined closely enough, will exhibit some nonlinearity. Consequently, the assumption of linearity is always an approximation. For many purposes, however, the results so obtained are quite useful. In any case, the relative ease with which linear systems can be handled mathematically provides considerable incentive for the use of linear descriptions when they are at all appropriate. That only the mathematical model and not the actual physical system is linear, however, should not be forgotten.

2. SYSTEM ELEMENTS

Most linear elements found commonly in feedback-control systems can be characterized by ordinary linear differential equations having real constant coefficients. Thus, if the box of Figure 1 represents an element of some sort -- electric motor,

Figure 1 - System Element

*Read at meeting of Study Group on Feedback Control Systems Nov. 30, 1951.

hydraulic motor, mechanical or electrical network, or whatever -- the output x_o, a function of time t, often can be related to its input $x_i(t)$ by a differential equation of the form

$$a_n \frac{d^n x_o}{dt^n} + \cdots + a_1 \frac{dx_o}{dt} + a_o x_o = b_m \frac{d^m x_i}{dt} + \cdots + b_1 \frac{dx_i}{dt} + b_o x_i \qquad (1)$$

in which the coefficients a_i and b_i are real and constant.

3. EXAMPLE OF MOTOR

As an example, consider the d.c. electric motor represented schematically in Figure 2. The motor is driven by applying a

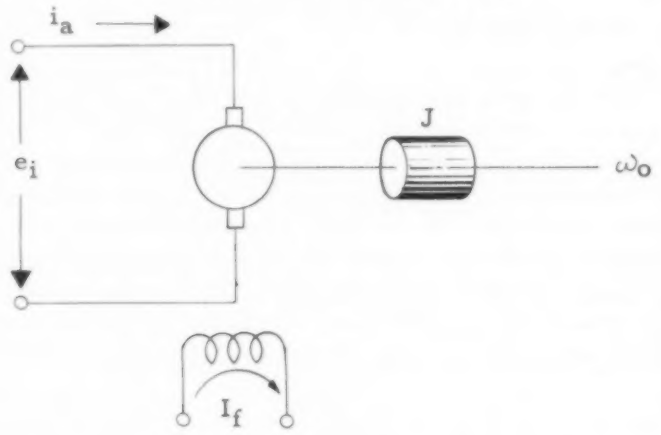


Figure 2 - Armature-Controlled d.c. Motor

voltage e_i to the armature terminals. The field current I_f is assumed to be held constant so that the torque developed by the motor is roughly proportional to the armature current:

$$\tau = k i_a \qquad (2)$$

Neglecting inductance of the armature circuit, i_a is given by

$$R_a i_a = e_i - e_m \qquad (3)$$

in which R_a is the armature resistance, and e_m is the back emf of the motor. The back emf in turn is proportional to motor speed

$$e_m = k\omega_o \tag{4}$$

The torque developed by the motor is absorbed by the load, which is assumed to be purely inertial, J being the inertia. These relations lead to the differential equation

$$J\frac{d\omega_o}{dt} + \frac{k^2}{R_a}\omega_o = \frac{k}{R_a}e_i \tag{5}$$

Equations of this type frequently are normalized to make them more generally applicable by defining the new constants T (called the "time constant") and K:

$$T = \frac{JR_a}{k^2} \tag{6}$$

$$K = \frac{1}{k} \tag{7}$$

The equation then may be written

$$T\frac{d\omega_o}{dt} + \omega_o = Ke_i \tag{8}$$

4. RESPONSE OF MOTOR TO STEP FUNCTION

If the form of the input voltage e_i is specified as a function of time the resulting speed ω_o can be found by solving the differential equation (8). Suppose, for example, that e_i has the form shown in Figure 3, i.e., a "step function" of magnitude V. For $t > 0$, (8) becomes

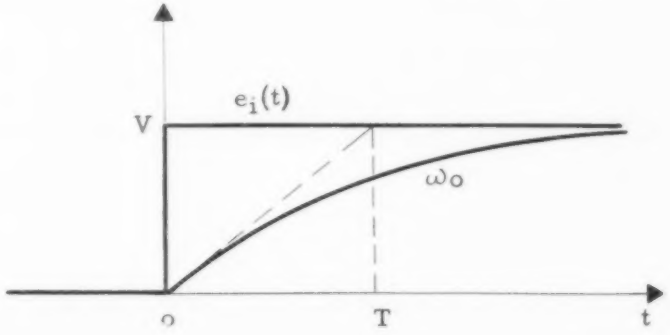


Figure 3 - Input and Response of Motor

$$T \frac{d\omega_o}{dt} + \omega_o = KV \tag{9}$$

The solution of linear differential equations having constant coefficients, such as (9), is the sum of terms of the form e^{pt} which satisfy the homogeneous equation (right member of the equation set equal to zero) and a particular solution of the inhomogeneous equation.

In (9), a particular solution is the steady-state term KV. The homogeneous equation is

$$T \frac{d\omega_o}{dt} + \omega_o = 0. \tag{10}$$

Assume

$$\omega_o = C\, e^{pt} \tag{11}$$

in which C is any constant. Substitution of (11) into (10) gives

$$T\, Cpe^{pt} + C\, e^{pt} = 0 \tag{12}$$

or

$$(Tp + 1)\, C\, e^{pt} = 0 \tag{13}$$

Since e^{pt} is never zero, (11) is a solution for nonzero C only if

$$Tp + 1 = 0 \tag{14}$$

or

$$p = -\frac{1}{T} \tag{15}$$

The general solution of (9) then is

$$\omega_o = Ce^{-\frac{t}{T}} + KV \tag{16}$$

If ω_o is initially zero, C must be -KV, so that in this case

$$\omega_o = KV\left(1 - e^{-\frac{t}{T}}\right) \tag{17}$$

The form of this solution is sketched in Figure 3.

5. TRANSFER FUNCTIONS

In terms of differential operator $D = \frac{d}{dt}$, (8) can be written

$$(TD + 1)\omega_o = Ke_i \tag{18}$$

or

$$\frac{\omega_o}{e_i} = \frac{K}{TD + 1} \tag{19}$$

The function

$$\frac{K}{TD + 1} = H(D) \tag{20}$$

is called the "transfer function" of the element. The transfer function of the more general element, described by (1), similarly is

$$H(D) = \frac{B(D)}{A(D)} \tag{21}$$

in which

$$A(D) = a_n D^n + \cdots + a_1 D + a_o \tag{22}$$

$$B(D) = b_m D^m + \cdots + b_1 D + b_o \tag{23}$$

6. SOLUTION OF THE GENERAL EQUATION

If the input $x_i(t)$ to the element of Figure 1 is given, the right side of (1) is a known function of time. Here again the output $x_o(t)$ is the sum of a particular integral together with terms of the form e^{pt} which satisfy

$$A(D)\, x_o = 0 \tag{24}$$

With e^{pt} substituted in (24) for x_o, the result is

$$A(p)\, e^{pt} = 0 \tag{25}$$

from which

$$A(p) = 0 \tag{26}$$

If A is an n th degree polynomial there are n values of p, p_k, which satisfy (26), called the characteristic equation. If the p_k are all distinct, the general solution of the homogeneous equation (24) is

$$x_o = \sum_{k=1}^{n} C_k e^{pkt} \tag{27}$$

in which the C_k are n arbitrary constants.

7. ROOTS OF THE CHARACTERISTIC EQUATION

The p_k in general are complex, say

$$p_k = \alpha_k + i\beta_k \tag{28}$$

But since the coefficients in A(D) are real, the p_k always occur in conjugate pairs so that the resulting solution can be real -- of the form

$$C_k e^{\alpha_k t} \cos (\beta_k t + \theta_k) \tag{29}$$

The values of p_k determine to a large extent the nature of the behavior of the element. In (29), for instance, if any $\alpha_k > 0$, an oscillatory mode is possible in which the amplitude of oscillation increases exponentially with time. On the other hand, if $\alpha_k < 0$, the amplitude decreases exponentially with time. For $\alpha_k = 0$, the amplitude remains constant. When $\beta_k = 0$, the mode is not oscillatory but merely a simple exponential, increasing or decreasing depending on whether $\alpha_k > 0$ or $\alpha_k < 0$.

In Figure 4, some possible values of p_k are spotted in the complex p-plane. The behavior of the modes corresponding to the various p_k is sketched nearby. Thus, a certain type of behavior is associated with each position in the p-plane.

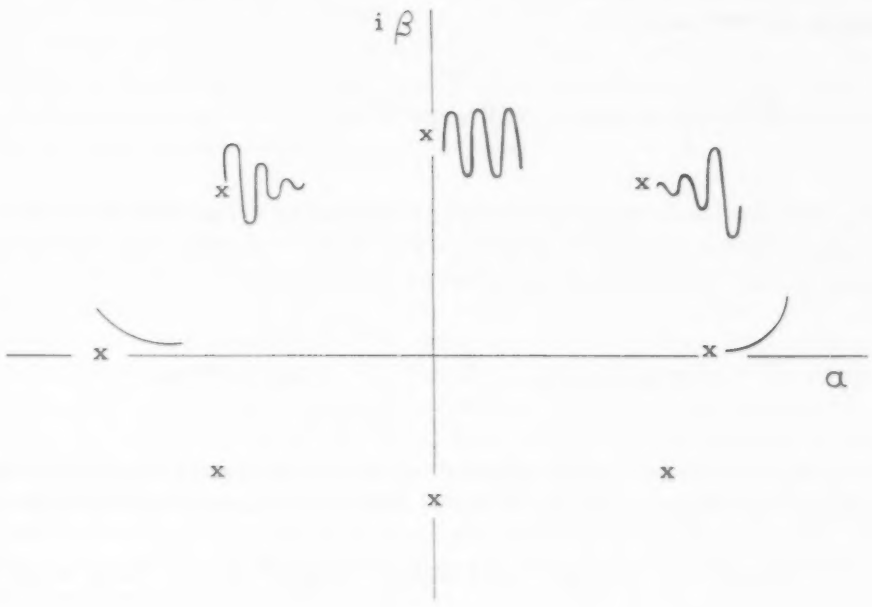


Figure 4 - Behavior Associated With Various Portions of the p-plane

8. STABILITY

A rigorous, completely satisfactory definition of "stability" is rather complicated and depends somewhat upon the situation. The general idea, however, is that in a stable element or system the effect of disturbances eventually dies out. Evidently an element having excitable modes represented by roots in the right half p-plane or on the imaginary axis would be unstable in this sense.

Much emphasis is placed on the question of stability in the analysis and design of feedback-control systems since an unstable system ordinarily is completely unsatisfactory. In fact simply to know whether a system is stable or not is not usually enough -- it is also important to know how stable a device is. More is said about absolute and relative stability later.

9. OPERATIONAL OR TRANSFORM METHODS

The response of the motor to a step function is obtained in Section 4 by the so-called classical solution of the differential equation. Operational methods based on the Fourier or Laplace transform also are used extensively in servo analysis.

The Laplace transform of a time function x(t) often is written and defined as follows:

$$\mathcal{L}\left[x(t)\right] = X(s) = \int_0^\infty x(t)e^{-st}\,dt \tag{30}$$

Thus, a function of time x(t) is transformed to a function X(s) of a complex variable by the integral of (30). The time function can be recovered by means of the corresponding inverse transform:

$$\mathcal{L}^{-1}\left[X(s)\right] = x(t) = \frac{1}{2\pi i}\int_{\alpha - i\infty}^{\alpha + i\infty} X(s)\,e^{st}\,ds \tag{31}$$

It follows from these definitions that the transform of (1), assuming the time functions and their first few derivatives are zero initially, is

$$(a_n s^n + \cdots + a_1 s + a_o)\,X_o(s) = (b_m s^m + \cdots + b_1 s + b_o)\,X_i(s) \tag{32}$$

or

$$A(s)\,X_o(s) = B(s)\,X_i(s) \tag{33}$$

In the above, $X_i(s)$ and $X_o(s)$ are the Laplace transform of the input and output respectively. If xi(t) is known, its transform $X_i(s)$ can be found either from the integral definition (30) or, as usually is the case, simply from previously prepared tables which list various pairs of corresponding t and s functions. Then the transform of the output is

$$X_o(s) = \frac{B(s)}{A(s)}\,X_i(s) = H(s)\,X_i(s) \tag{34}$$

from which the output as a function of time can be found by taking the inverse transform:

$$x_o(t) = \mathcal{L}^{-1}\left[X_o(s)\right] = \mathcal{L}^{-1}\left[H(s)\,X_i(s)\right] \tag{35}$$

If the initial values had not all been zero, another term would have appeared on the right side of (32) and (33), which would have been taken into account in evaluating the inverse transform. In this lies one of the chief differences between the solution of equations by Laplace transform methods and by classical methods. The classical approach yields a general

solution with n arbitrary constants -- the C_k of (27) -- which must be evaluated from the initial conditions. In the Laplace transform solution, on the other hand, initial conditions are taken into account in the transformation of the differential equation. The resulting time function eventually obtained through the inverse transformation, then, is already adjusted for the proper initial conditions.

Both methods of solution require factoring the characteristic equation to find the n p-roots, and both, of course, require a knowledge of the initial conditions. The classical method, as already mentioned, requires the evaluation of n constants, whereas the transform method often requires a partial-fraction expansion or the equivalent. The transform method seems to be somewhat more systematic and may provide a neater way of obtaining the particular integral for some input forcing functions. Other reasons for its extensive use are that some useful results of a general nature are easily obtained in this way and that transform methods aid in unifying transient and sinusoidal steady-state concepts. For the actual computation of the solution to a specific equation, however, the classical method often is more direct and just as tractable as transform methods, particularly for low order equations.

The transfer function of an element frequently is written as a function of s and considered to be a Laplace transform. Thus, the transfer function of the electric motor, Figure 2, may be written

$$H(s) = \frac{K}{Ts + 1} \tag{36}$$

[Compare with (20).]

10. THE SINUSOIDAL STEADY STATE

Section 4 considered the result of applying a step function command to the electric motor. This type of input is used frequently as a test signal to provide an indication of the performance of an element or system. Another standard test signal is the sinusoid. Suppose, for example, that the motor input $e_i(t)$ is sinusoidal, say

$$e_i(t) = E_i \cos \omega t = \mathrm{Re}\, E_i\, e^{i\omega t} \tag{37}$$

Because of the nature of the differential equation (8) governing the motor, the steady state response also is sinusoidal. It may be written as

$$\omega_o(t) = \mathrm{Re}\, \Omega_o\, e^{i\omega t} \tag{38}$$

in which Ω_o may be complex. Substitution of this assumed form of the solution into the differential equation (8) gives

$$\mathrm{Re}(Ti\,\omega\Omega_o\, e^{i\omega t} + \Omega_o\, e^{i\,\omega t}) = K\,\mathrm{Re}\,E_i\, e^{i\,\omega t} \tag{39}$$

Since this must hold for any t, it reduces to

$$(i\,\omega\,T + 1)\,\Omega_o = KE_i \tag{40}$$

or

$$\frac{\Omega_o}{E_i} = \frac{K}{i\omega T + 1} = H(i\omega).$$

Thus the ratio of the complex amplitudes of output and input is simply the transfer function, now written as a function of $i\,\omega$. In this form the transfer function sometimes is called the sinusoidal or frequency response function.

11. FEEDBACK-CONTROL SYSTEMS

Control systems usually are composed of several elements, each having its own transfer function. A typical servo might be represented schematically as shown in Figure 5. In addition to the differencing element which forms the error θ_e from the input θ_i and output θ_o, the system contains three elements with the transfer function $H_1(s)$, $H_2(s)$, and $H_3(s)$. The over-all transfer function of the forward part

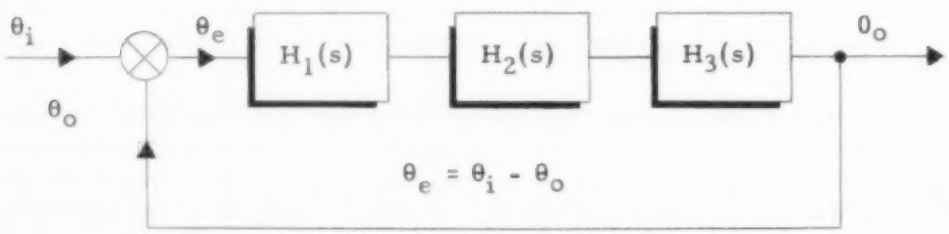


Figure 5 - Schematic Representation of a Servo

of the loop, relating θ_o to θ_e, is the product of the individual transfer functions:

$$\frac{\theta_o(s)}{\theta_e(s)} = H_1(s)\, H_2(s)\, H_3(s) = H_o(s) \tag{41}$$

The transform of the error equation is

$$\theta_e(s) = \theta_i(s) - \theta_o(s) \tag{42}$$

The transfer function which describes the behavior of the complete system is obtained from (41) and (42) by eliminating $\theta_e(s)$:

$$\frac{\theta_o(s)}{\theta_i(s)} = \frac{H_o(s)}{1 + H_o(s)} \tag{43}$$

12. SYSTEM STABILITY

In Sections 7 and 8 we saw that the stability of an element depends upon the position in the complex plane of the zeros of the denominator of the transfer function. For the system of Figure 5, these zeros are the roots of

$$1 + H_o(s) = 0 \tag{44}$$

or

$$H_o(s) = -1 \tag{45}$$

Therefore, in order that the complete system be stable, the roots of (44) or (45) must lie in the left half of the complex plane.

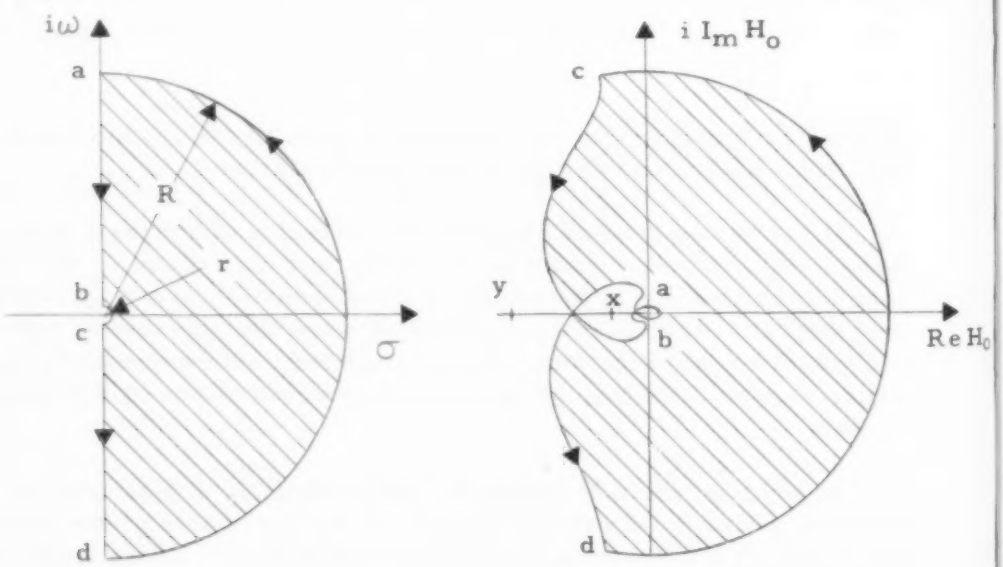
If the characteristic equation can be factored and the actual roots obtained, the stability question is easily settled. Usually this would be a long and tedious job if the characteristic equation is a polynomial of very high degree. And since the relation between the polynomial zeros and the individual system elements is somewhat indirect, a complete redetermination of the zeros may be necessary even for slight modifications of the system.

A simpler method might be application of a test such as Routh or Hurwitz, which determines the signs of the real parts of the zeros without the necessity for factoring a high order polynomial. These tests, however, give directly no measure of the relative stability of the system. And here again, slight modification of the system may require completely new calculations.

13. THE NYQUIST CRITERION

By far the most useful tool for the design of feedback systems at the present time is the Nyquist stability criterion with its variations and extensions. Use of this criterion involves plotting the open-loop transfer function $H_o(s)$ in the complex plane as s traces out a closed contour which includes the right half s-plane. This contour in the s-plane usually has the form sketched in Figure 6a which, for large enough R and small enough r, enclose all of the right half-plane that is of interest. As s traverses the contour of Figure 6a, $H_o(s)$ takes on various values which when plotted produce a closed contour or locus in the $H_o(s)$ plane. A typical locus is sketched in Figure 6b. Corresponding values along the s and H_o loci are labeled. Thus, the positive imaginary axis in the s-plane, from large ω to small ω, leads to the portion of the H_o locus from a to b. The small semicircular detour around the origin in the s-plane is taken to avoid the point $s = 0$ at which the $H_o(s)$ for most servo systems becomes infinite.

The Nyquist criterion states that, if the individual elements whose transfer functions make up $H_o(s)$ are themselves stable, the complete system is stable provided the $H_o(s)$ locus does not enclose the -1 point. If, for example, the scale of the sketch of Figure 6b is such that the point x corresponds to -1, the Nyquist criterion indicates that the system is unstable. However, if y corresponds to -1, the system is stable.



(a) s plane (b) $H_o(s)$ plane

Figure 6

As a result of the fact that the coefficients in differential equations describing the system are real, the H_o locus for ω negative is always the mirror image about the real axis of the locus for positive ω. Consequently, to apply the Nyquist criterion to the usual servo, it is actually necessary to plot only the transfer function $H_o(i\omega)$ as ω varies from very large to very small positive values. For stability, then the -1 point must lie to the left of the locus.

14. RELATIVE STABILITY

One of the great advantages of this approach is that an idea of the relative stability of a system can be obtained simply by noticing how closely the $H_o(i\omega)$ locus approaches the -1 point. For reasonable system stability, it turns out that the locus should not enter a circle in roughly the position shown in Figure 7. A locus plot such as this, then, gives a clue as to how the elements of the loop should be designed so that when the loop is closed the system will perform satisfactorily. A little experience shows, for example, what modifications or additions might be made to a system possessing locus a, Figure 7, so as to produce the more acceptable locus b.

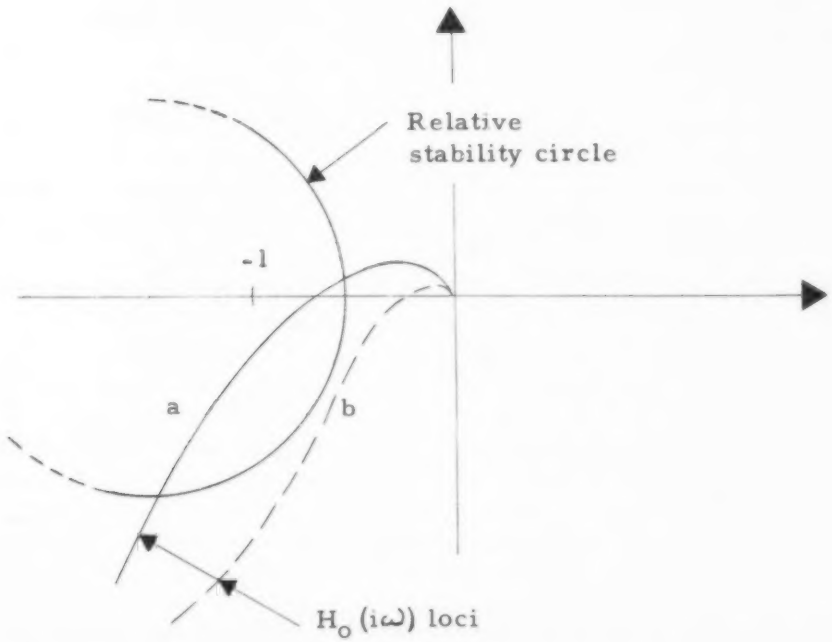


Figure 7

15. THE LOGARITHMIC TRANSFORMATION

A variation in this procedure that is of considerable practical value consists of plotting the logarithm of $H_o(i\omega)$ rather than the function itself. The log of $H_o(i\omega)$ is

$$\ln H_o(i\omega) = \ln \left|H_o(i\omega)\right| + i \text{ Ang } H_o(i\omega)$$

It is customary to plot the log of the magnitude along the vertical axis in units of $20 \log_{10} \left|H_o(i\omega)\right|$ (called decibels whether associated directly with power or not) and the angle along the horizontal axis, as in Figure 8. The point 0 decibels, -180° corresponds to the -1 point of Figure 7, and the relative stability circle there transforms to the oval sketched in Figure 8. A typical locus also is sketched in Figure 8.

This transformation simplifies design procedures primarily because the magnitudes of the individual transfer functions can be

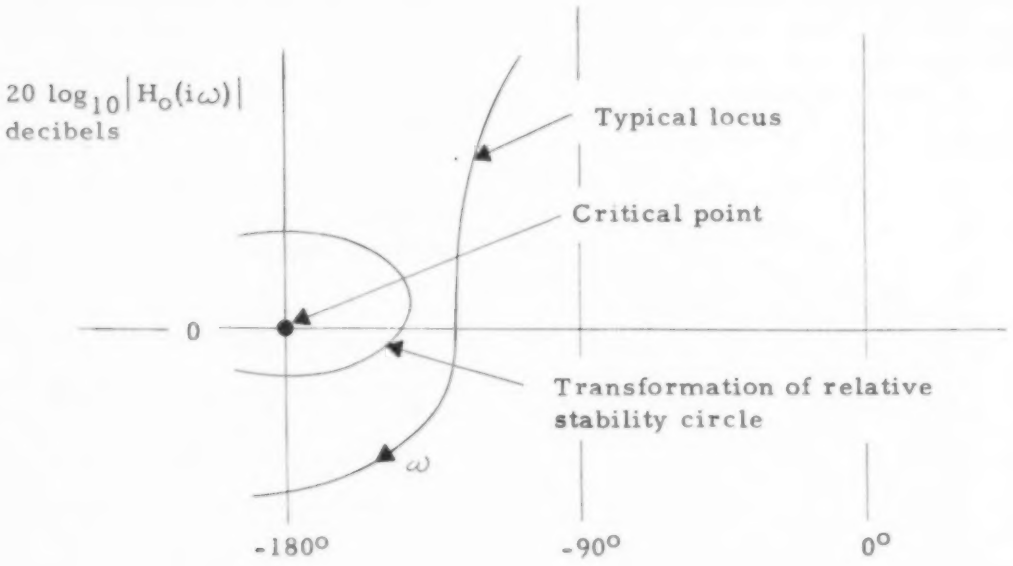


Figure 8

multiplied to obtain the over-all transfer function H_o simply by adding their logarithms. Also, there exist simple graphical constructions for logarithms of the magnitudes of the individual transfer functions.

ILLUSTRATIONS

Figure		Page
1.	Fixed control action at 60 ft/sec. Checkerboard test with step input at steering wheel	66
2.	Free control action Schematic diagram of observations with impulse disturbance at steering wheel	69
3.	Block diagram of car control	69
4.	Attitude of car in transient turn	73
5.	Relations between slip angles at front, center, and rear	73

DIRECTIONAL CONTROL OF AUTOMOBILES*

Robert Schilling
Research Laboratories Division, General Motors Corporation

ABSTRACT

The paper describes theoretical and experimental approaches to the problem of directional stability of automobiles. For a specific passenger vehicle, the lowest mode of oscillation has calculated frequency in the order of .7 cps. Damping is about .5 of critical under conditions of fixed control and about .25 under free control. These values are confirmed approximately by experiment. Frequency and damping of the second mode are only roughly predicted by calculation.

Whenever men have tried to construct a new type of vehicle such as the airplane or the automobile, it has been quite clear to them from the beginning that some kind of directional control is necessary, but only gradually have they realized that controllability alone is not enough; the vehicle must also possess something generally called "directional stability." By this term is usually meant the ability to follow a desired path without directional oscillations. Efforts to produce such a property by cut and try methods alone have often been unsatisfactory and some understanding of the fundamental dynamics, often combined with considerable analytical work, has been necessary.

The introduction of softer tires and softer suspensions of the independent type on American passenger cars led to such radical changes in the directional properties that research into the fundamentals of steering was begun. Valuable information was developed, but even now the problem is by no means fully understood.

The directional control forces of a vehicle come either from internal effects, such as inertia or gyroscopic forces, or from the interaction of the vehicle and the medium in which it travels, such as hydraulic forces on a vessel, aerodynamic forces on an aircraft, and ground contact forces on a road vehicle. These ground contact forces of pneumatic tires were investigated by Evans[1] and by Bull[2]. Their effect on the car's action in curve equilibrium conditions was discussed by Olley[3, 4] and was measured

---

* Read at the regular meeting of the society, Feb. 12, 1953.

by the so-called "skidpad test" at the General Motors Proving Ground.[5] This test furnishes a great deal of information, but since it is a static or equilibrium test, it cannot measure the dynamic properties of the car which appear in transition maneuvers. A dynamic transition test, the "checkerboard test," was also developed, using the rather cumbersome instrumentation techniques available at the time.

Recorded data from one such test are shown in Figure 1, which is a reproduction of Figure 19 in reference 5. This illustration shows at the bottom the path of the car over the checkerboard, proceeding from left to right and next above this, the steering angles measured at the front wheels and at the steering wheel. It can be seen that the maneuver consisted in subjecting the car to a step function input at the steering wheel, first to the right and then to the left. The upper traces record slip and roll angle and finally path angles and lateral accelerations, all as functions of distance traveled. Note that it required approximately 60 ft of travel or one second's time for the car to reach equilibrium in roll, slip angles, and lateral acceleration. The yaw acceleration appears as the difference between front and rear acceleration; it can be seen to change sign during the transition, indicating considerably less than critical yaw damping. A second mode of oscillation with a frequency of approximately 2 cps is also evident in the slip angle and path angle records.

These test data show conclusively that the directional control and stability of a car is very similar to that of a boat or an aircraft. In all these cases, the directional control process appears as an oscillating system of at least two degrees of freedom, that is, as a problem of engineering mechanics. It is therefore of interest to see if equations can be developed which describe this action and which are based on known properties of tires, suspension, steering system, and the car's inertia. The Appendix shows that such an analysis in its simplest linearized form leads to a fourth order differential equation, in which one variable, here the yaw velocity, is treated as the dependent variable, while the steering angle is used as the independent variable. Numerical values, corresponding to a representative medium size car, are used to compute frequency and damping of the two modes of free oscillation. These numerical values, however, were not measured on the test car itself, but are based on measurements on various similar cars.

The conditions of Case 1 in Table I correspond to the test run in Figure 1 and show fair agreement. The lower mode has a calculated frequency of .72 cps which is somewhat slower than that shown by the test; its specific damping of .69 seems to fit the test quite well. The higher mode with a calculated frequency of 1.07 cps and a specific damping of .38 does

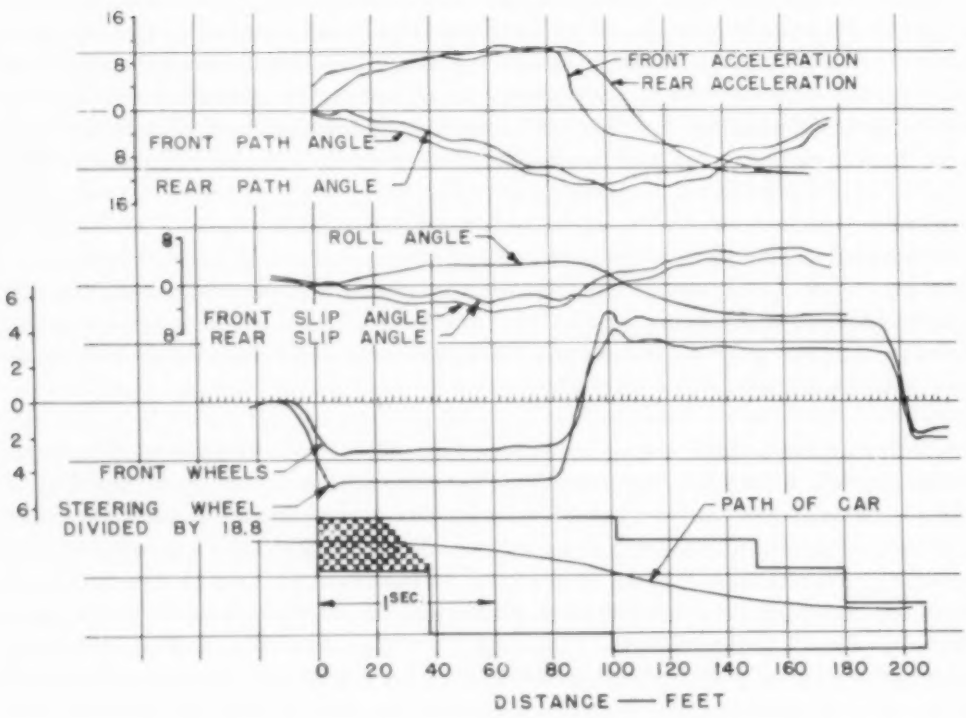


Figure 1 - Fixed Control Action at 60 ft/sec.
Checkerboard Test with Step Input at Steering Wheel

not correspond to the higher mode of about 2.0 cps indicated by the test. This may be due to a wrong choice of the numerical values, but it is also possible that this apparent 2 cps motion is not real. It may simply represent scatter of test data as stated in the original paper[5]. The two calculated modes of .69 cps and 1.07 cps can easily combine so that in this test they could not be separated.

Comparison of Case 1 and Case 2 shows that doubling the speed has only a minor effect on the two modes for this car, so that a driver would probably not detect any change. This is usually true for well-behaved cars.

Both the test and the analysis fall short of expressing directional stability in a satisfactory way for two reasons. They do not show the disturbance produced by a road wave or a gust of wind. This shortcoming could fairly easily be corrected by an extension of test and analysis.

Besides this, and more fundamental, is the objection that a driver's impression is based very much on steering wheel "feel"; that means on the reaction forces he receives at the steering wheel during a maneuver and on a wavy or windy road. We know that changes in caster and in the reversibility of the steering mechanism can greatly change what the driver calls directional stability. The checkerboard test and the analysis of the Appendix are concerned with the properties of the car under "fixed control" conditions, and hence contain no information on steering reaction forces.

Some of the desired information can be obtained by running a checkerboard test under "free control." This consists in throwing the car off balance by jerking the steering wheel and then releasing it. The car will then oscillate about a straight line while the steering wheel swings back and forth. Such a maneuver is depicted in Figure 2. The frequency and decay of the car's oscillation can be observed without much instrumentation; they correlate well with some items of steering performance in normal driving - for example, with action on an S curve and with recovery from a partial skid. Such tests have been run on the checkerboard, but no completely processed data are available.

The analysis can also be modified to simulate roughly this type of control. If the front wheels caster freely, with zero friction in the steering linkage, their cornering forces must be zero at all times; hence, setting the front wheel cornering power equal to zero, we get a first approximation of the free control stability equation.

Case 3 of the Appendix shows that the calculation for this condition leads to a first mode with a frequency of .78 cps and .34 specific damping.

Observation on the road usually shows frequencies higher than this, sometimes as high as 2 cps; however, the analysis correctly predicts a much lower specific damping for free control than for fixed control. Comparison of Case 3 and Case 4 shows that the damping of both modes decreases materially with increasing speed, which also checks with road observation. Some cars will actually show a build-up of this oscillation at high car speed, indicating negative damping. However, change of static friction in the steering linkage usually affects this free control oscillation markedly, and it would be a mistake to draw too many detailed conclusions from this simplified analysis.

Even with the addition of the free control test and free control analysis, we still lack some of the data required to specify what the driver calls "handling" of a car. We still need to know what forces appear at the steering wheel if the driver fully or partially prevents the wheel from moving in any arbitrary maneuver. After we know this, we will of course ask just how the driver reacts to the car's motions and forces; we will want to know just what items in this collection of data really correspond to his impression of directional stability and of responsiveness. As a means of finding this final correlation, our past approach is not very promising.

We will probably have to use techniques of servo motor analysis by investigating a block diagram such as Figure 3, according to the suggestion of Paul F. Chenea.[6] The input here is a disturbance produced by the road or the wind, the upper row of blocks representing transfer functions of the car which establish the values of significant variables produced by the directional properties of the car. The driver is represented by the lower block which contains his transfer function or functions; he picks up signals from some variables, acts on them, and produces as output a new steering angle. This is added to the original disturbance, thus forming a closed loop system.

At present we do not know enough about the human operator as a servo to write down the transfer function we need in this diagram; it is even possible that our present mathematics is incapable of giving us a satisfactory function. But research now in process on human reaction and on suitable mathematics may soon give us the proper tools. We shall then be able to consider the stability of the whole system of car and driver so that we can specify the directional properties of the car which will make it pleasant and safe even in the hands of a poor driver.

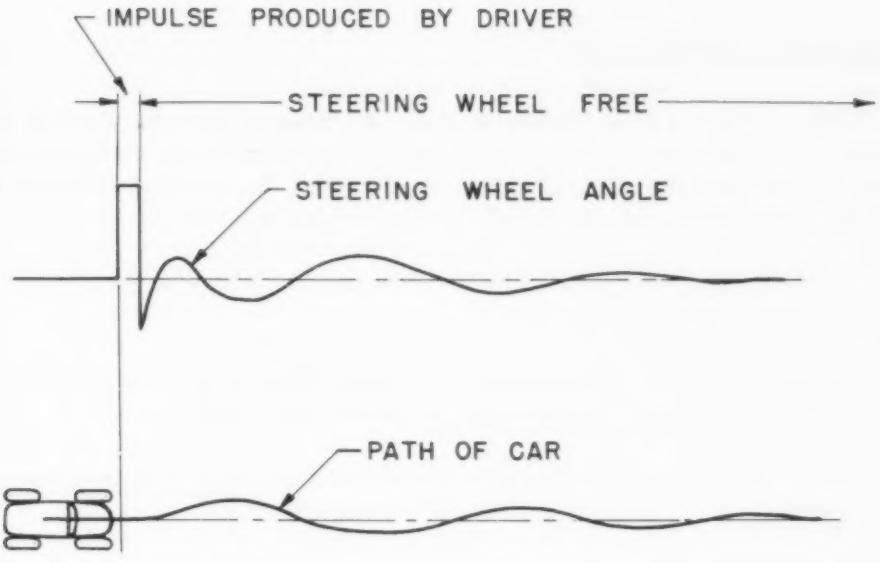


Figure 2 - Free Control Action
Schematic Diagram of Observations with Impulse Disturbance at Steering Wheel

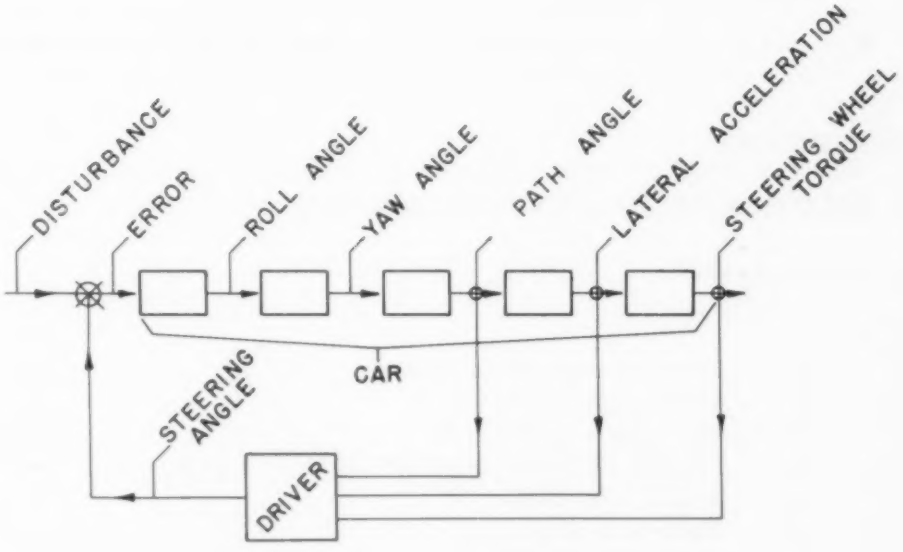


Figure 3 - Block Diagram of Car Control

ACKNOWLEDGEMENTS

The author is indebted to Mr. K. A. Stonex for permission to reproduce his illustration, to Prof. Paul F. Chenea for permission to use extracts from his lecture notes, and to Mr. L. G. Johnson for carrying out the numerical calculations for the Appendix.

BIBLIOGRAPHY

1. Evans, R. D., "Properties of Tires Affecting Riding, Steering, and Handling," SAE Transactions, February 1935.

2. Bull, A. W., "Tire Behavior in Steering," SAE Transactions, August 1939.

3. Olley, M., "Independent Suspension - Its Whys and Wherefores," SAE Journal, vol. 34, p. 73, 1934.

4. Olley, M., "National Influences on American Passenger Car Design," Proceedings IAE, vol. XXXII, p. 509.

5. Stonex, Kenneth A., "Car Control Factors and Their Measurements," SAE Transactions, March 1941.

6. Chenea, Paul F., Purdue University, unpublished lecture notes.

APPENDIX

SOLUTION OF EQUATIONS OF MOTION

List of Symbols

ϕ = Course angle (yaw angle)
ψ = Path angle
γ = Steering angle
α_o = Center slip angle (referred to roll axis)
α_1 = Front slip angle
α_2 = Rear slip angle
θ = Roll angle
F = Lateral inertia force at C.G.
F_1 = Front cornering force
F_2 = Rear cornering force

m = Mass of car
I_x = Moment of inertia, around an axis parallel to the roll axis - through the C.G.
I_z = Moment of inertia, around yaw axis
L = Wheelbase
a = Horizontal distance from C.G. to front wheels
b = Horizontal distance from C.G. to rear wheels
h = Distance of C.G. above roll axis
V = Car speed
λ = Lateral acceleration at C.G.
C_1 = Front cornering power
C_2 = Rear cornering power
r_1 = Front roll-steer ratio
r_2 = Rear roll-steer ratio
K = Roll spring rate about the roll axis
f = Roll damping coefficient about the roll axis

According to references 1 and 2 there is a linear relation between slip angle and cornering force for slip angles below six degrees, so that we may write

$$F = C \times \alpha$$

The test data also show a variation of cornering power with vertical load; this will be neglected in order to avoid non-linearity. To the slip angles produced by the tire must be added steering angles due to lateral load in the suspension. These will simply be included by modification of the cornering power C.

Additional steering angles are produced by roll displacement and will be assumed proportional to roll angle. This leads to

$$\alpha_1 = \frac{F_1}{C_1} + r_1 \theta$$

$$\alpha_2 = \frac{F_2}{C_2} + r_2 \theta$$

or, solving for F:

$$F_1 = C_1 (\alpha_1 - r_1 \theta)$$

$$F_2 = C_2 (\alpha_2 - r_2 \theta)$$

From Figure 4 we can then write three equations of equilibrium.

Lateral forces:

$$m\lambda - C_1 (\alpha_1 - r_1 \theta) - C_2 (\alpha_2 - r_2 \theta) = 0 \qquad (1)$$

Moments about yaw axis:

$$I_z \ddot{\Phi} - C_1 a(\alpha_1 - r_1 \theta) + C_2 b(\alpha_2 - r_2 \theta) = 0 \qquad (2)$$

Moments about roll axis:

$$I_x \ddot{\theta} + f \dot{\theta} + K \theta - mh \lambda = 0 \qquad (3)$$

Additional equations are derived from geometric relations based on Figure 5.

$$\dot{\Phi} = \frac{V_1 \sin(\gamma - \alpha_1) + V_2 \sin \alpha_2}{L}$$

$$= \frac{V_1 \sin (\gamma - \alpha_1) + V_o \sin \alpha_o}{a}$$

$$= \frac{V_2 \sin \alpha_2 - V_o \sin \alpha_o}{b}$$

For small angles where $\alpha_1 = \sin \alpha_1$ and $V_1 = V_2 = V_o = V$, these three equations are:

$$\dot{\Phi} = \frac{V}{L} (\gamma - \alpha_1 + \alpha_2) \qquad (4)$$

$$= \frac{V}{a} (\gamma - \alpha_1 + \alpha_o) \qquad (4b)$$

$$= \frac{V}{b} (\alpha_2 - \alpha_o) \qquad (4c)$$

From (4b) and (4c) $\alpha_o = (1/L) \left[a\alpha_2 - b (\gamma - \alpha_1) \right]$ Hence,

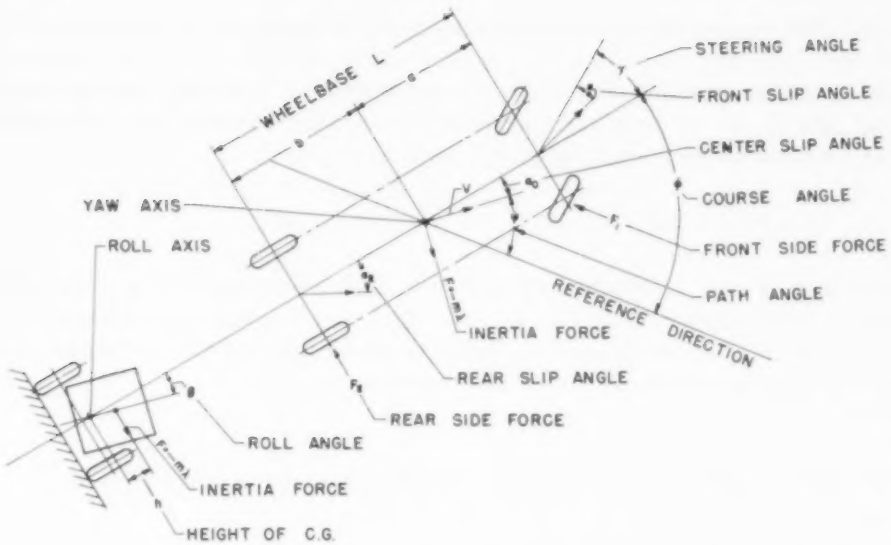


Figure 4 - Attitude of Car in Transient Turn

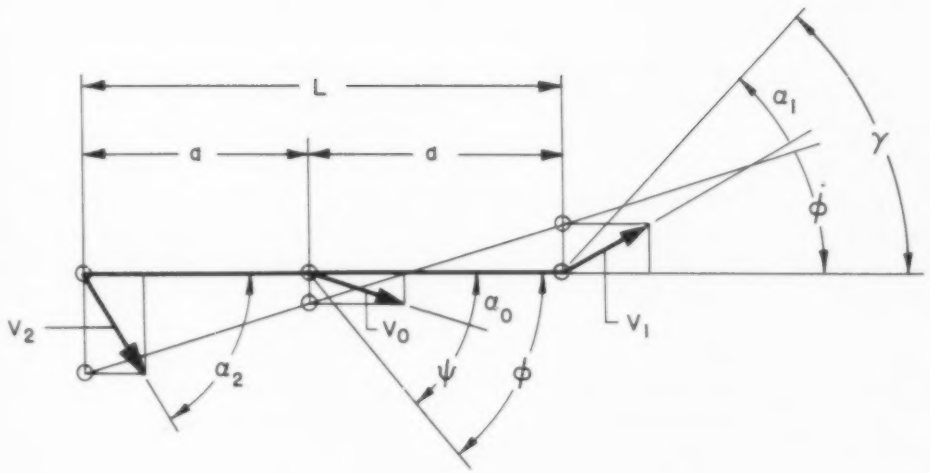


Figure 5 - Relations Between Slip Angles at Front, Center, and Rear

$$\psi = \Phi - \alpha_o = \Phi - (1/L)\left[a\alpha_2 - b(\gamma - \alpha_1)\right] \tag{5}$$

A last equation results from the consideration that the lateral acceleration at the C.G. is the sum of lateral acceleration at the roll axis plus lateral acceleration relative to the roll axis.

$$\lambda = V\dot{\psi} - h\ddot{\theta} \tag{6}$$

Considering V as constant, γ as the independent variable, we have six dependent variables ψ, Φ, Θ, λ, α_1, α_2, so the six equations (1) to (6) are sufficient. We may solve them for one dependent variable, for example the yaw velocity Φ, and find the equation

$$\Phi^V + a_4\Phi^{IV} + a_3\dddot{\Phi} + a_2\ddot{\Phi} + a_1\dot{\Phi} = b_3\dddot{\gamma} + b_2\ddot{\gamma} + b_1\dot{\gamma} + b_o\gamma \tag{7}$$

Where the parameters have the values:

$$a_4 = \frac{f}{I_x} + \frac{(C_1 + C_2)(1 + \frac{mh^2}{I_x})}{mV} + \frac{C_1a^2 + C_2b^2}{I_zV}$$

$$a_3 = \frac{f(C_1 + C_2)}{m\,I_x\,V} + \frac{h(C_1r_1 + C_2r_2)}{I_xI_z} + \frac{K}{I_x} + \frac{C_1C_2L^2(1 + \frac{mh^2}{I_x})}{m\,I_z\,V^2}$$

$$+ \frac{f(C_1a^2 + C_2b^2)}{I_xI_z\,V} + \frac{C_2b - C_1a}{I_x}$$

$$a_2 = \frac{K(C_1 + C_2)}{m\,I_x\,V} + \frac{C_1C_2Lf}{m\,I_xI_z\,V^2} + \frac{C_1C_2hL(ar_2 + br_1)}{I_xI_z\,V} + \frac{K(C_1a^2 + C_2b^2)}{I_xI_z\,V}$$

$$+ \frac{f(C_2b - C_1a)}{I_xI_z}$$

$$a_1 = \frac{C_1C_2KL^2}{m\,I_xI_z\,V^2} + \frac{C_1C_2hL(r_1 - r_2)}{I_xI_z} + \frac{K(C_2b - C_1a)}{I_xI_z}$$

No complete set of numerical measurements for any one car is available, but values for a representative medium sized car, partly based on measurements and partly on estimates, would be as follows:

mg	= 4100 lb
m	= 127.3 lb sec^2/ft
I_z	= 3000 lb ft sec^2
I_x	= 400 lb ft sec^2
L	= 10 ft
a	= 4.8 ft
b	= 5.2 ft
h	= 1.7 ft
C_1	= 14,000 lb/rad
C_2	= 14,000 lb/rad
k	= 34,500 lb ft/rad
f	= 1300 ft lb sec/rad
r_1	= +.15
r_2	= -.10

With these values and car speeds of V = 60 ft/sec (41 mph) and V = 120 ft/sec (82 mph), the values of the parameters are for fixed control (i.e., $C_1 = C_2 = 14,000$ lb/rad)

	At 60 ft/sec	At 120 ft/sec
a_4	= 14.18	8.72
a_3	= 140.06	107.25
a_2	= 676.69	340.22
a_1	= 2085	1163.

For free control (i.e., $C_1 = 0$, $C_2 = 14,000$ lb/rad

	At 60 ft/sec	At 120 ft/sec
a_4	= 8.87	6.06
a_3	= 123.31	116.92
a_2	= 418.35	248.61
a_1	= 2093	2093.

Hence the characteristic equations of equation (7) are for

FIXED CONTROL: (i.e., $C_1 = C_2 = 14,000$ lb/rad)

Case (1) V = 60 ft/sec

$D^4 + 14.18\ D^3 + 140.06\ D^2 + 676.69\ D + 2085 = 0$

Factors are: $(D^2 + 8.69D + 39.328)(D^2 + 5.49D + 53.024) = 0$

Roots are: $-4.345 \pm 4.5221i$ and $-2.745 \pm 6.7446i$

Frequencies are: .72 cps and 1.07 cps

Specific dampings are: .69 and .38

Case (2) V = 120 ft/sec

$D^4 + 8.72\ D^3 + 107.25\ D^2 + 340.22\ D + 1163 = 0$

Factors are: $(D^2 + 3.527D + 15.936)(D^2 + 5.193D + 72.999) =$

Roots are: $-1.7635 \pm 3.5814i$ and $-2.5965 \pm 8.1399i$

Frequencies are: .57 cps and 1.30 cps

Specific dampings are: .44 and .30

FREE CONTROL: (i.e., $C_1 = 0$, $C_2 = 14,000$ lb/rad)

Case (3) V = 60 ft/sec

$D^4 + 8.87\ D^3 + 123.31\ D^2 + 418.35\ D + 2093 = 0$

Factors are: $(D^2 + 3.545D + 27.042)(D^2 + 5.325D + 77.390) =$

Roots are: $-1.7725 \pm 4.8888i$ and $-2.6625 \pm 8.3846i$

Frequencies are: .78 cps and 1.33 cps

Specific dampings are: .34 and .30

Case (4) V = 120 ft/sec

$D^4 + 6.06\ D^3 + 116.92\ D^2 + 248.61\ D + 2093 = 0$

Factors are: $(D^2 + 1.646D + 24.609)(D^2 + 4.414D + 85.045) = 0$

Roots are: $-.823 \pm 4.8920i$ and $-2.207 \pm 8.9540i$

Frequencies are: .78 cps and 1.43 cps

Specific dampings are: .17 and .24

TABLE I

Case	Condition on		1st Mode		2nd Mode	
	Control	Car Vel. fps	Freq. cps	Spec. Damping	Freq. cps	Spec. Damping
1	Fixed	60	.72	.69	1.07	.38
2	"	120	.57	.44	1.30	.30
3	Free	60	.78	.34	1.33	.30
4	"	120	.78	.17	1.43	.24

ILLUSTRATIONS

Figure		Page
1.	Straight Tooth Generation	81
2.	Curved Tooth Generation	82
3.	Principle of Bevel Gear Generator	84
4.	Cradle and Tool Slides in Straight Bevel Generator	85
5.	Bevel Gear Generator - for Straight Teeth	86
6.	Bevel Gear Generator - for Spiral Teeth	87
7.	Spiral Bevel Generator (with imaginary generating gear)	88
8.	Conjugate Racks	89
9.	Extension of Sang's Theory	90
10.	A Theoretical Method for Exactly Conjugate Spiral Bevel Gear Teeth	91
11.	Spherical Transverse Surface	94
12.	Transverse Layout of Straight Bevel Gear Teeth	96
13.	Layout of Spiral Bevel Gear Teeth in Normal Plane	97
14.	Tooth Bearing in Spiral Bevel Gears	99

(Photographs of Nos. 4, 7, and 14 by Courtesy of Gleason Works)

BEVEL GEARS
THEIR GENERATION, KINEMATICS, AND GEOMETRY*

Allan H. Candee, Rochester, New York

ABSTRACT

This tutorial paper outlines the geometry and fabrication of bevel gears with particular reference to the methods used at the Gleason Works, from which the author recently retired.

DEFINITION

In a pair of bevel gears we have two axes of rotation that intersect, and a pair of imaginary pitch cones that roll together and are a convenient reference for study and design. Usually, but not always, the velocity ratio is constant and the cones are circular. The common element of the pitch cones is an instant axis of relative rotation.

KINDS OF TEETH

The kinds of teeth made in practice do not conform to old conventional theories, but are the result of practical manufacturing processes.

We generate straight teeth by cutting tools that reciprocate on straight slides of a rotating machine cradle (Fig. 1). Relative to the cradle, the straight cutting edges move in planes that usually pass through the apex point of the bevel gear. Then the relative motion in the machine results in the generation of tooth surfaces the profiles of which are not exactly involute but do make fully conjugate gears. In this operation the tools and the machine represent an imaginary "generating gear."

To generate spiral teeth, which are curved instead of straight, a tool moves in a curved cutting path, which similarly partakes of the rotation of the cradle (Figs. 2 and 6). In the spiral bevel generators used in the automotive industry, and also elsewhere, the tool rotates about a cutter axis; but there are other generating machines in which the tool moves in a curved path that does not have an axis.

---

* Condensation of paper read at the regular meeting of the Society October 14, 1953.

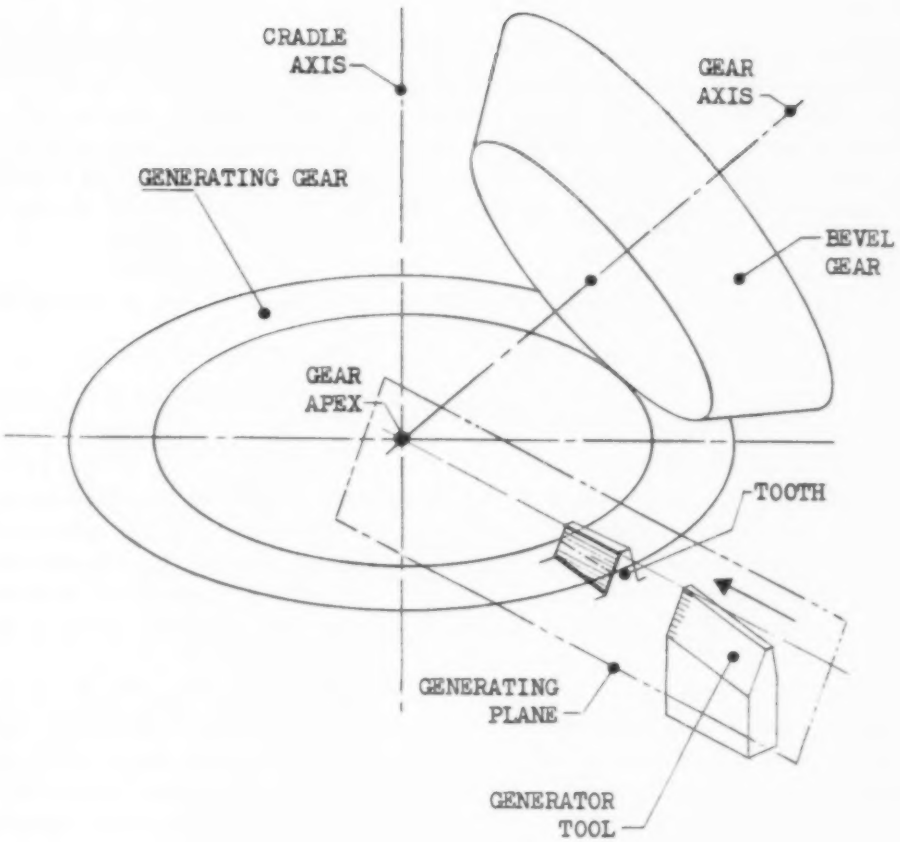


Figure 1 - Straight Tooth Generation

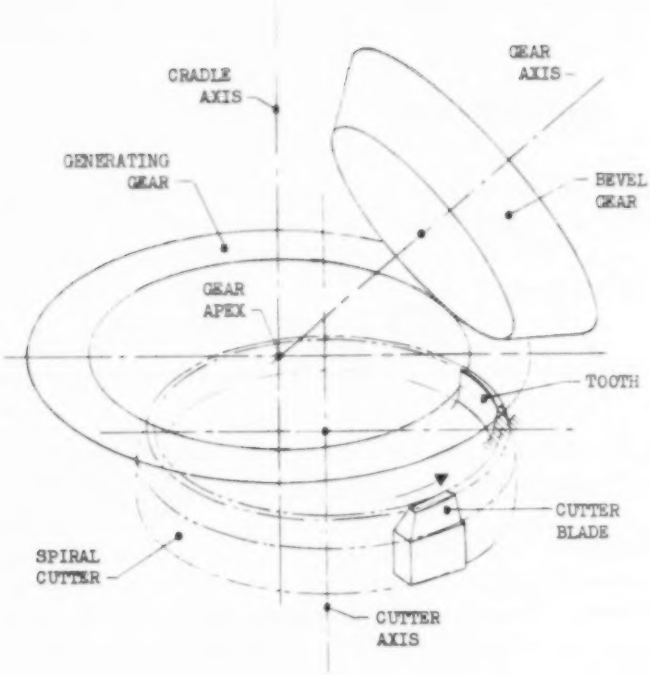


Figure 2 - Curved Tooth Generation

GENERATING MACHINES

The principle of operation of a straight bevel gear generator is shown in Figs. 3 to 5. There is a cradle carrying two reciprocating tools, each to generate one side of a tooth. A gear segment (Fig. 3) or a gear train provides the constant "ratio of roll" between cradle and gear blank, which rotate about intersecting axes. The tool slides are set to a "tooth angle" to produce the required taper on tooth thickness (Fig. 5).

The tool tips travel in a plane perpendicular to the cradle axis, which is the root plane of the bevel gear. The instant axis makes a small angle with this plane. Thus, the "generating gear" has less than a 90-degree pitch angle, which makes the profile curvature slightly greater than that generated by an exact crown gear. Some newer straight bevel gear generators allow adjustments for any pitch angle of the imaginary "generating gear."

Figures 6 and 7 show the principle of a spiral bevel gear generator.

CONJUGATE GEAR TEETH AND INTERCHANGEABLE GEARS

Any tooth profile or lengthwise tooth formation of conjugate racks yields conjugate gears in the sense that for each point of one tooth surface there is only one mating contact point on the mating surface. Sang's[1] theory, developed a hundred years ago, emphasized the symmetry of the paths of contact on either side of the rack pitch line and for opposite sides of the teeth as a condition to obtain interchangeability.

Fig. 8 shows the conjugate racks for interchangeable spur gears. Fig. 9 shows conjugate gears in the form of an internal and a counterpart external generating gear for conjugate spur gears; and also shows the generating gears of conjugate bevel gears. In this case, the pitch surfaces may be planes (for crown-gear generation) or matching cones, and the tooth surfaces can be of any convenient shape, unless interchangeability is required.

In many spiral bevel gears the generating surface of the cutter is a section of a cone. In the original conception, the axis of this cone (cutter axis in Fig. 10) is perpendicular to the pitch surface of the "generating gear." The "generating gear" lies in a plane tangent to the pitch cone at the instant axis. This type of cutter and its application was invented by James E. Gleason[2] some 40 years ago. In this case, the teeth are of constant depth. To obtain tapered tooth depth as in Fig. 6, certain de-

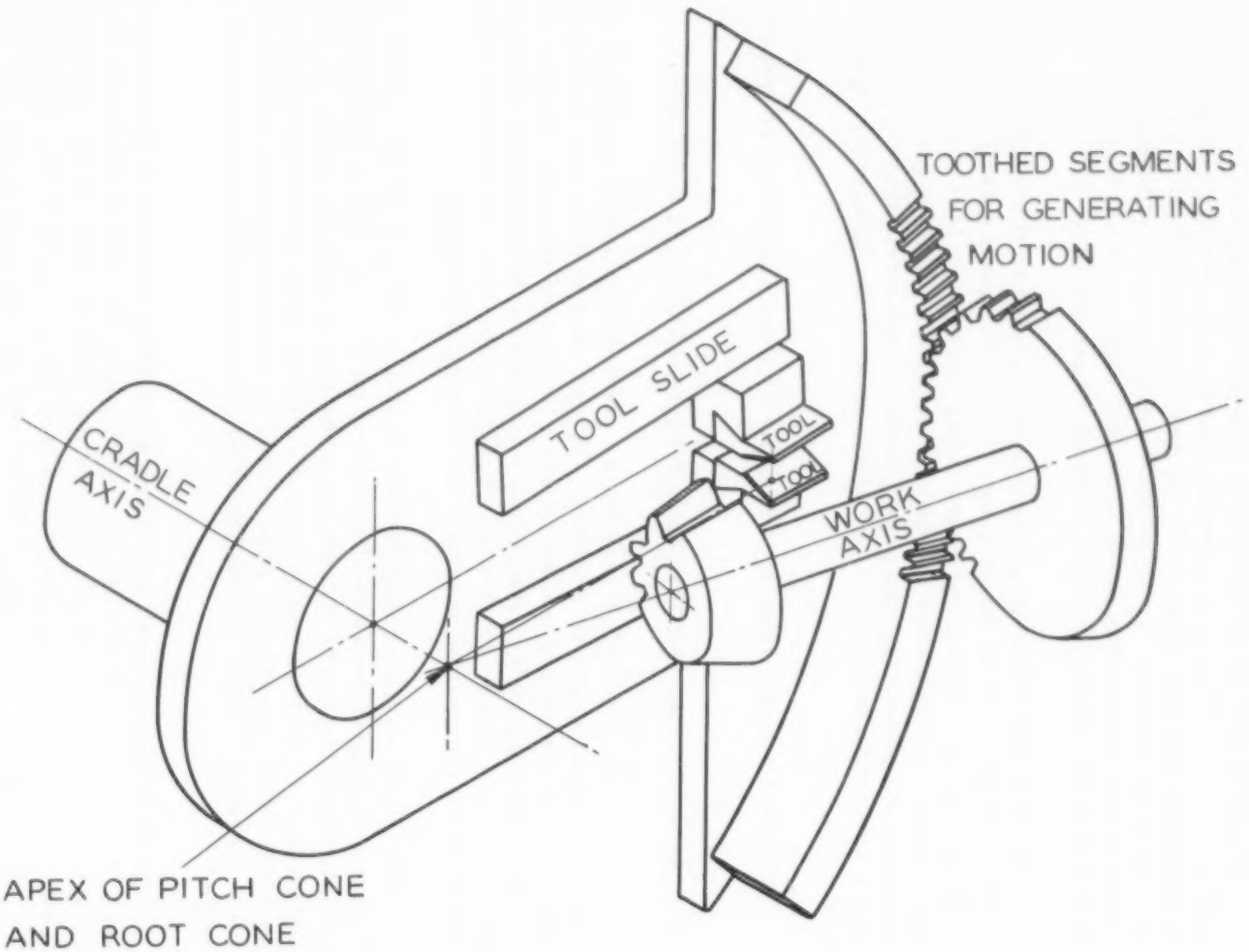


Figure 3 - Principle of Bevel Gear Generator

Figure 3 - Principle of Bevel Gear Generator

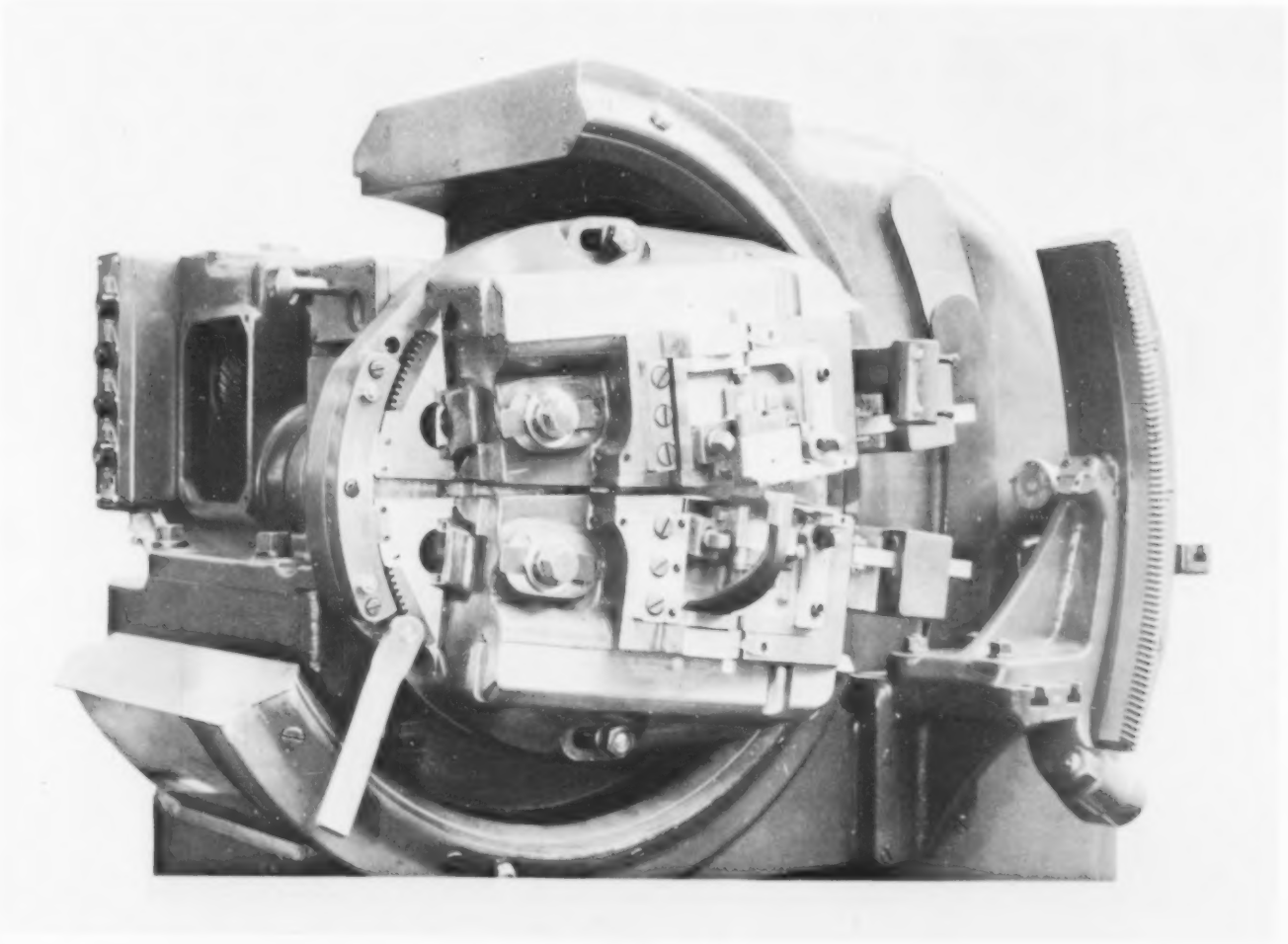


Figure 4 - Cradle and Tool Slides in Straight Bevel Generator

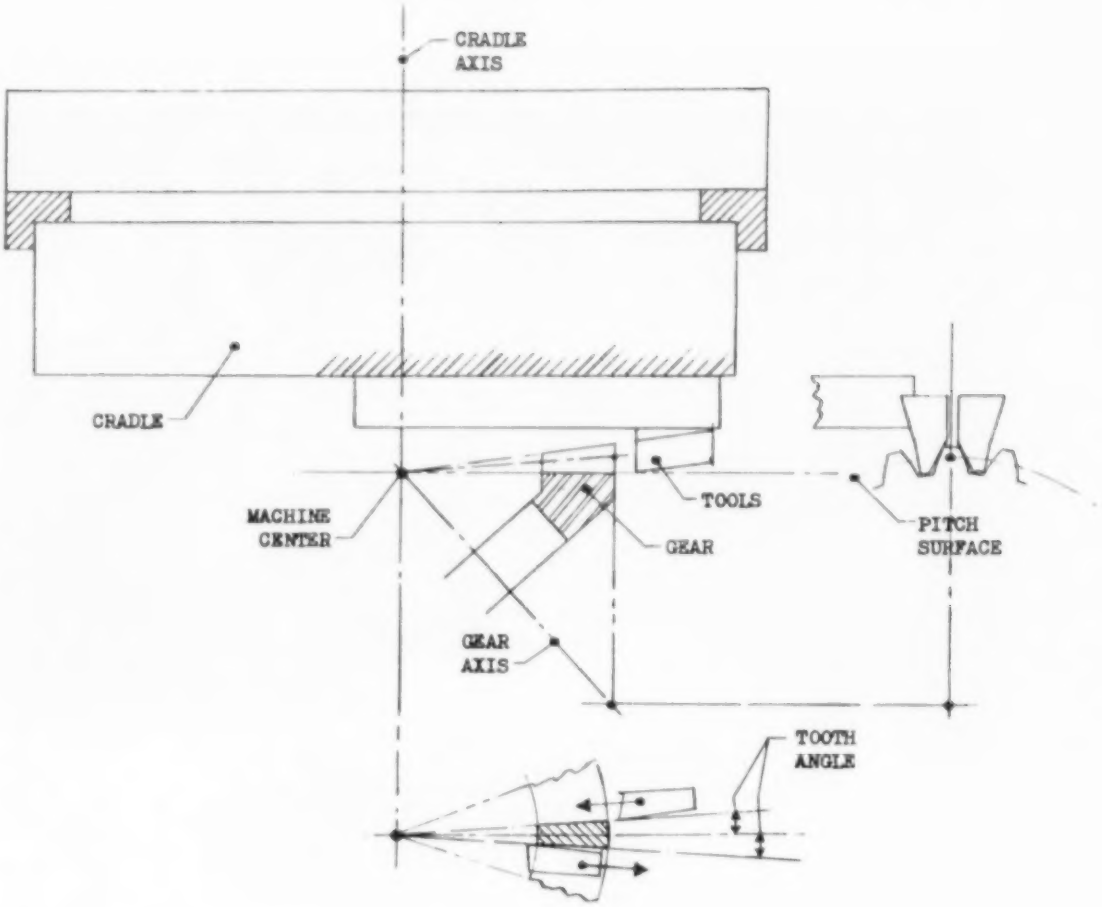


Figure 5 - Bevel Gear Generator - for Straight Teeth

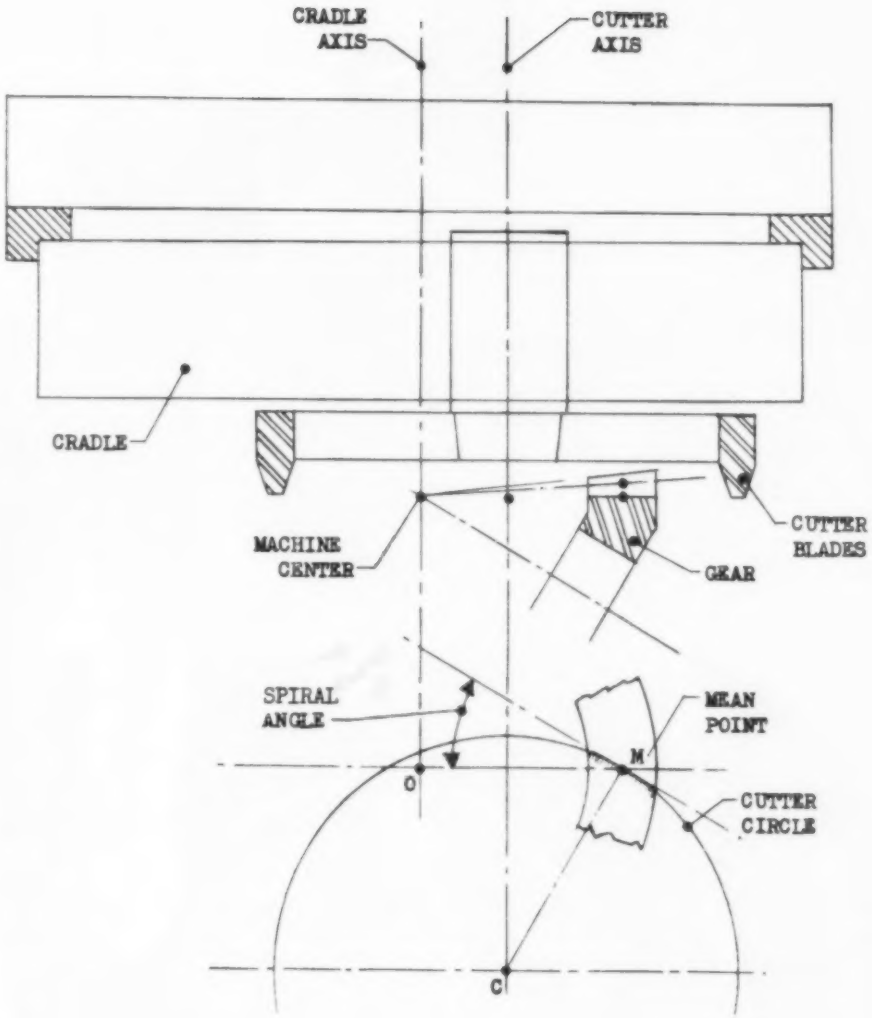


Figure 6 - Bevel Gear Generator - for Spiral Teeth

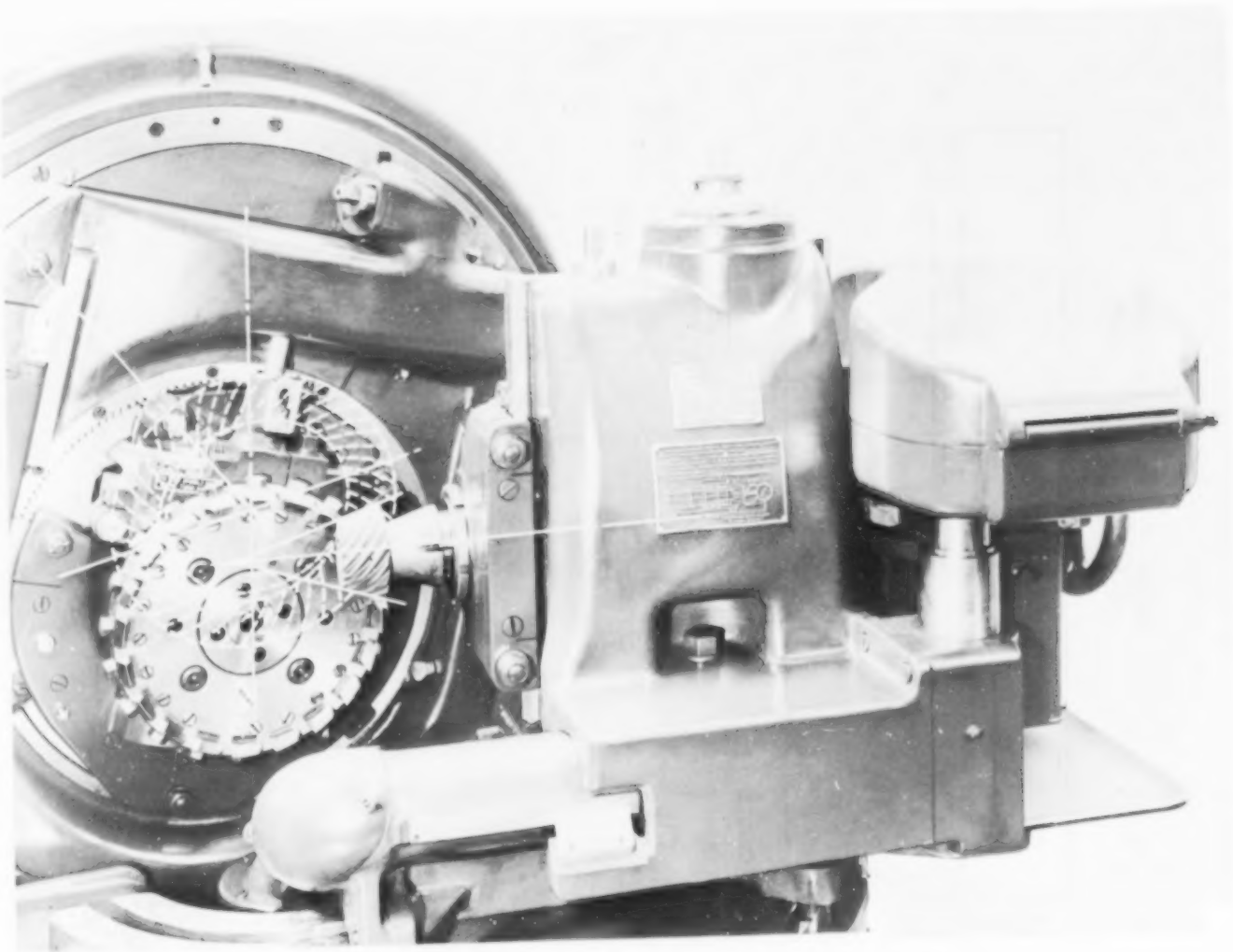


Figure 7 - Spiral Bevel Generator (with imaginary generating gear)

Figure 7 - Spiral Bevel Generator (with imaginary generating gear)

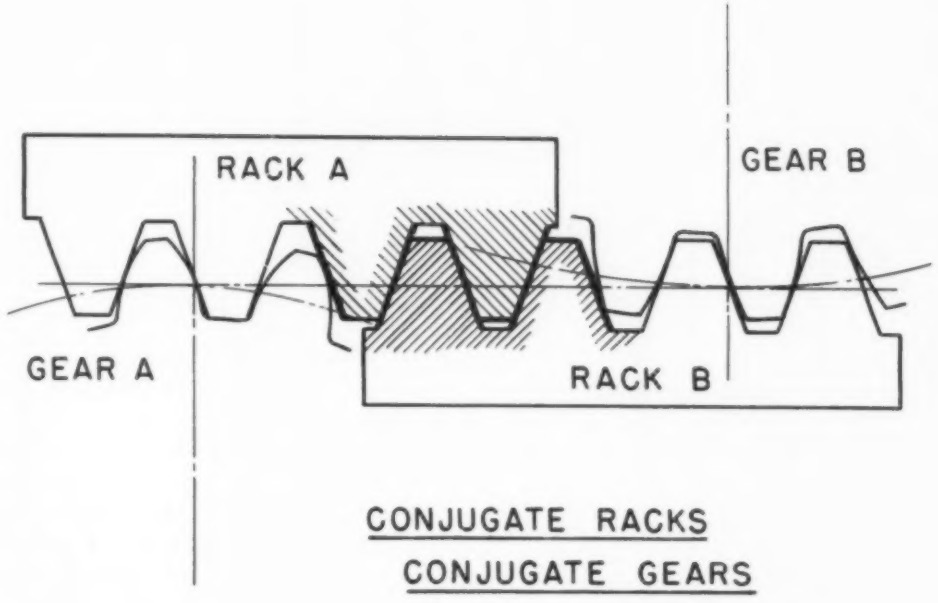


Figure 8 - Conjugate Racks

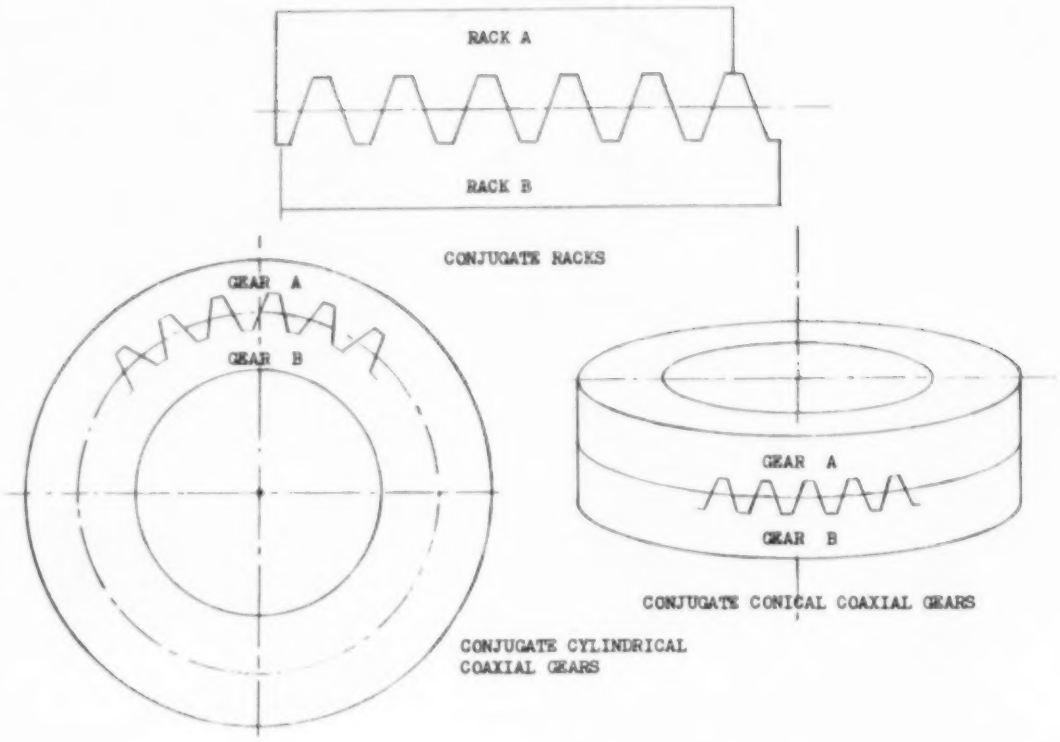


Figure 9 - Extension of Sang's Theory

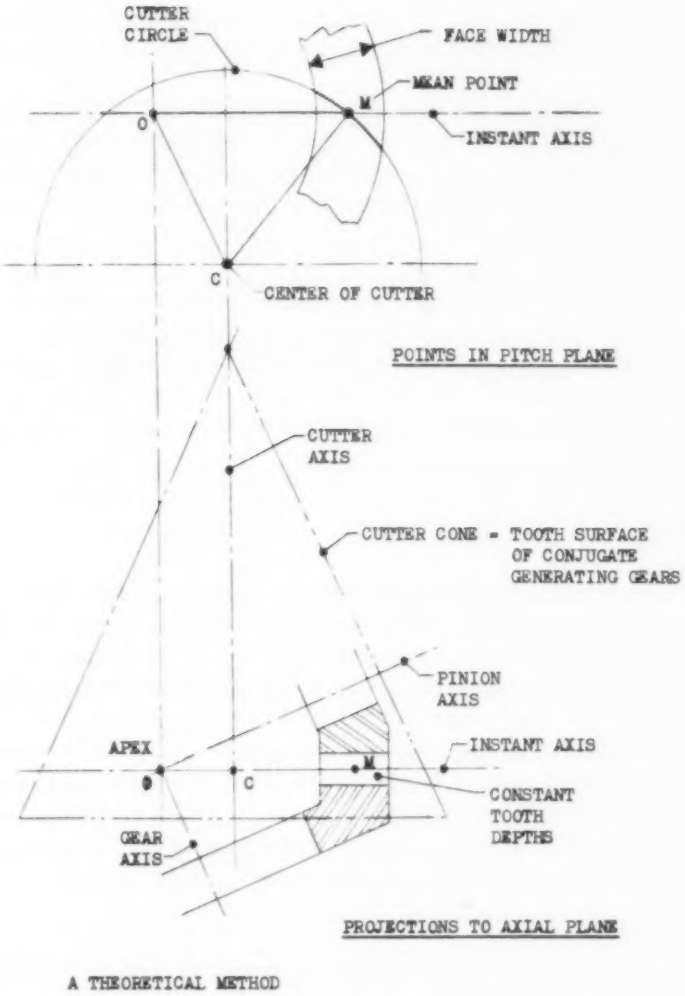


Figure 10 - A Theoretical Method of Exactly Conjugate Spiral Bevel Gear Teeth

partures from exact theory are made, for which there is a practical solution. However, exactly conjugate teeth in bevel gears are not desirable under operating conditions, for the same reasons which have led more recently to profile modification and crowning of cylindrical involute gears.

FORMATE† BEVEL GEARS

A characteristic of any gear-generating method is that the shape of the generating tool is not reproduced on the surface of the teeth of the gear. Forming methods of gear cutting do reproduce the tool shape, and such methods are generally faster and can be performed on more rigidly built machines.

For many years it has been more economical to make the larger member of a pair of spiral bevel or hypoid gears form-cut, rather than generated. Then the mating pinion is generated using a cutter similar to the form cut gear with which the pinion is to run. It is readily seen that a Formate gear and pinion can be made correctly conjugate in this way.

SURFACE CHANGES

Bevel gear generation permits more freedom than the processes in use for cylindrical gears. The tooth sides are often cut by separate tools or in separate operations, even in different machines, so that tooth thickness is not determined by tool dimension and cutting depth. Modifications of pressure angle and spiral angle, even when one tool is employed, are used to control tooth contact or bearing. An interesting example is the use of a "generating gear" with less than 90° pitch angle, to increase the profile curvature of the bevel gear by a few thousandths of an inch to modify the contact with the mating gear. Spur gears could also be modified in this manner if similar tools and methods were in use.

POINTS OF CONTACT, LINES OF CONTACT, AND CONJUGATE POINTS

Even though conjugate bevel gears can be produced without exact knowledge of tooth profiles and lengthwise curvature, such knowledge is often desired to estimate tooth strength and surface durability. The methods used are much the same as for spur gears and start with the determination of points of contact and lines of contact.

---

† Registered trade mark of the Gleason Works

A line of contact between two meshing teeth is the line or curve in which two curved surfaces contact each other at a given instant. When motion occurs, this line changes in location, sometimes also in shape.

A point of such a line can be found by the fundamental principle (common to spur and bevel gears but not to crossed axes gears) that the common normal at a point of contact passes through the instant axis.

It is therefore only necessary to drop normals from the instant axis to the (instantaneously fixed) tooth surface in order to establish points of such a line of contact. This procedure can be considerably simplified if the normals are dropped to the surface of the "generating gear," whose shape is geometrically simpler than that of the operating gear, e.g., plane in the crown gear of straight bevel gear teeth and conical in the crown gear of spiral bevel gear teeth. Lines of contact for some consecutive tooth positions determine the surface of action which, in spiral bevel gears, is a warped surface. Fortunately, we do not require these for general engineering purposes.

The surface of the "conjugate" generated gear tooth can be established by transferring points of the projections of the consecutive common contact lines to a selected fixed position of the required conjugate gear tooth.

This drawing procedure is cumbersome because it is necessary to establish three different projections of each contact point and each conjugate point. Locations of such points can be determined as accurately as desired by numerical calculations based on triangulation.

TOOTH LAYOUTS

The shape of a tooth section is necessary for estimating its strength. Cylindrical helical gears are not studied in a true section in a normal plane, but by substituting an equivalent spur gear as a satisfactorily close approximation. The equivalence is with respect to the size of the pitch circle, the circular pitch, circular tooth thickness, addendum and dedendum, pressure angle, and profile shape. Similar approximations are satisfactory for bevel gears.

TRANSVERSE SPHERE AND TRANSVERSE PLANE

Profile curves taken on a spherical surface centered at the apex point will contact one another in the same manner as the transverse profile of spur gears. Figure 11 shows such transverse sphere with pitch circles,

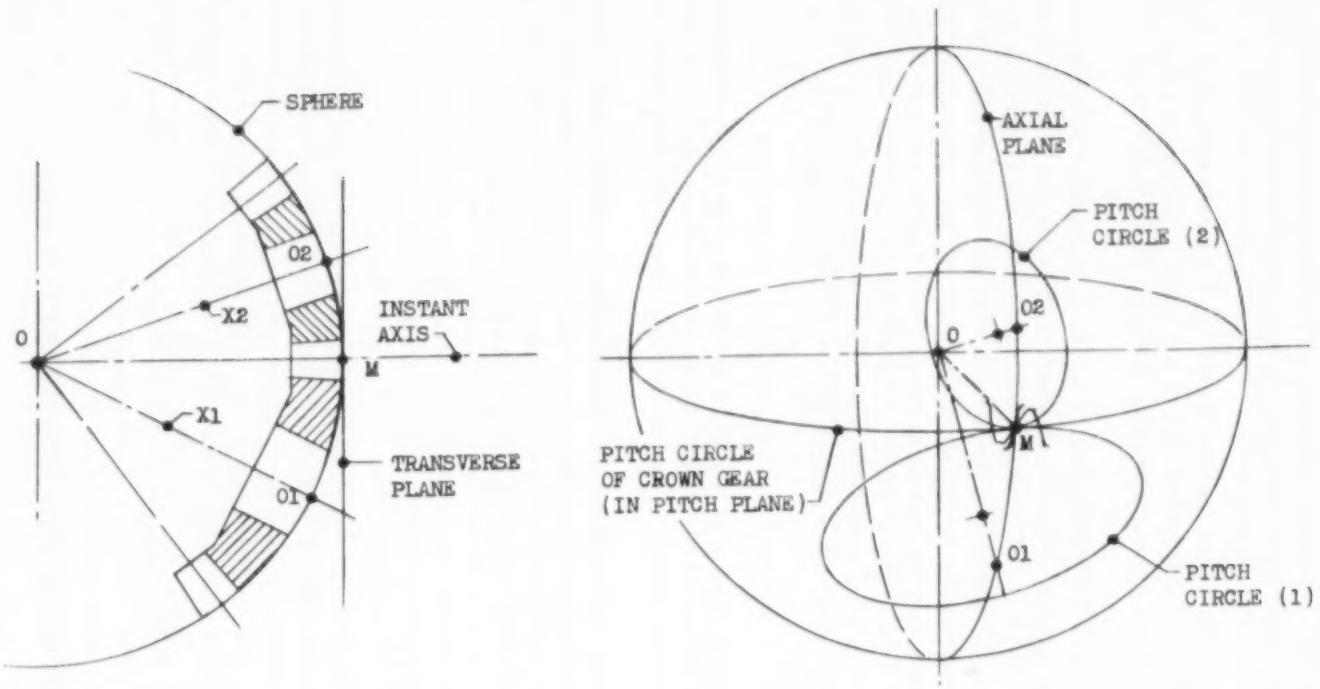


Figure 11 - Spherical Transverse Surface

profile curves, etc. Spherical geometry could be used to obtain exact results. Approximations based on plane line drawings are usually preferred.

For a straight bevel gear, Figure 12, an equivalent spur gear is substituted, drawn in a transverse plane perpendicular to a pitch cone element usually at the outer end of the tooth or, for engineering purposes, at the mean section. This method is known as Tredgold's[3] approximation and assumes the equivalent spur gear on the surface of a back cone unwrapped into a plane. Its pitch radius is the back cone distance or the normal distance of a point on the pitch cone element from the gear axis. It is also the radius of a sphere inscribed in the pitch cone. According to Meunier's [4] theorem for the curvature of surfaces, it is the radius of curvature of the pitch cone surface in the chosen transverse plane.

NORMAL LAYOUT OF SPIRAL TEETH

The pitch radius in the normal layout of a spiral bevel gear may be obtained by Euler's[4] theorem as the transverse radius divided by the square of the cosine of the spiral angle as graphically illustrated in Figure 13.

The approximate tooth profiles of the bevel gear are obtained by rolling the equivalent pitch circle of the generating gear with that of the bevel gear. The generating crown gear is represented as a rack with straight pitch line.

SUMMARY OF SPECIFICATIONS AND MACHINE SETTINGS

While approximations are adequate for tooth layouts, exact calculations must be made for cutter specifications and machine settings. Many different sizes and types of machines exist, and two or three generating methods may be used, even on the same machine. This does not allow standardization to the same degree as in spur and helical gears.

INSPECTION AND MEASUREMENT

The problem of bevel gear inspection is, in general, not approached by the same methods of exact measurement that are used for cylindrical involute gears, because the size and shape of the bevel gear tooth varies from toe to heel and because there is more than one way to set up a generating machine for gears of given numbers of teeth and pitch. Also, there is no simple way to follow the tooth profile reliably and accurately. The

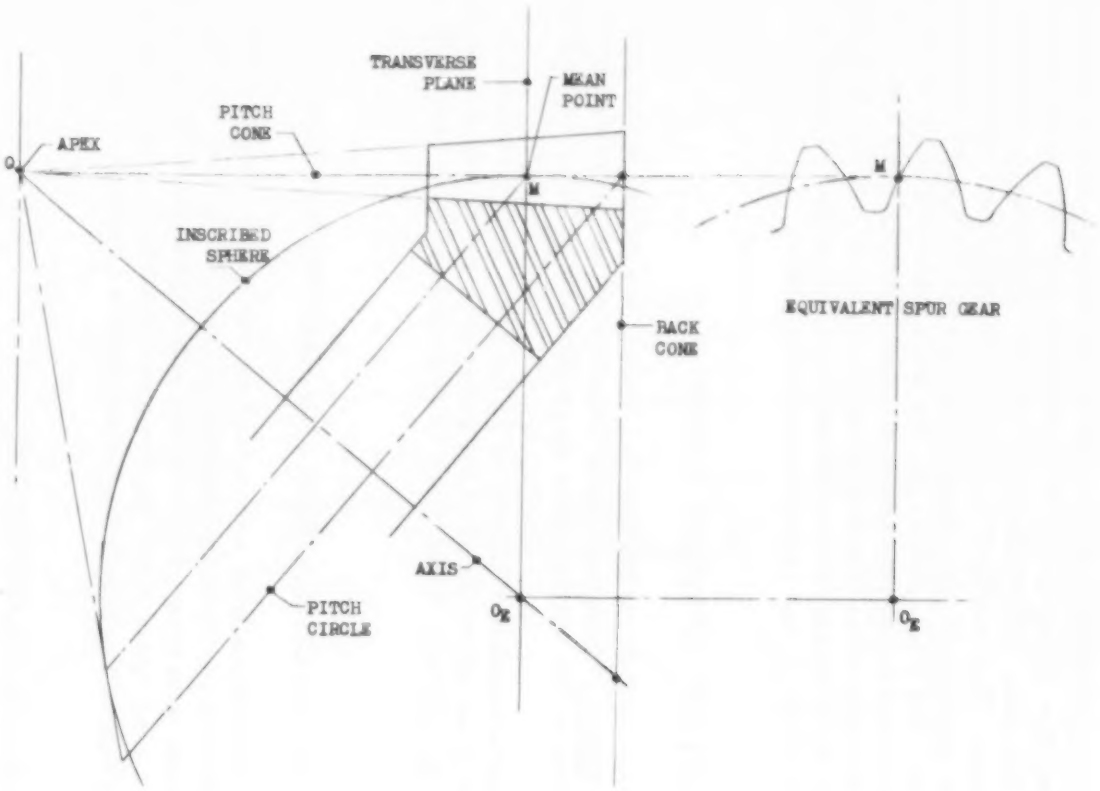


Figure 12 - Transverse Layout of Straight Bevel Gear Teeth

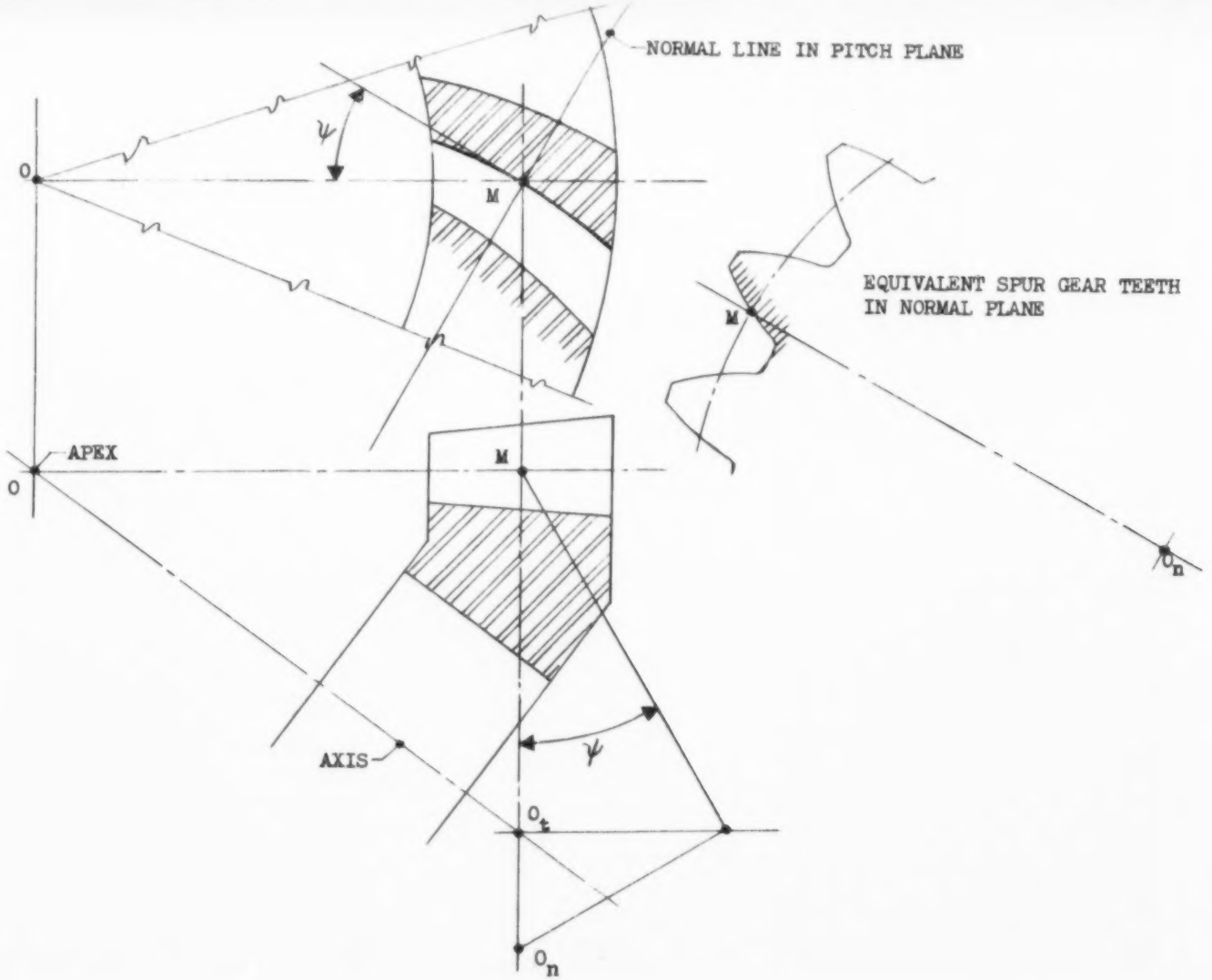


Figure 13 - Layout of Spiral Bevel Gear Teeth in Normal Plane

position at which a measurement is taken becomes as important as the measured dimensions; also the machine, the tools, the setup, and the selected generating motion affect the tooth thickness. There are more variables even than in worm gear teeth, which are not ordinarily inspected for tooth shape.

The real test of bevel gears consists of running a gear and pinion together and observing sound and tooth bearing (Fig. 14). The so-called "V-and-H check" is an example. In the central relative position the gears should contact with only a portion of the tooth surface. The tooth bearing can be shifted from one end of a tooth to the other by slight changes in the relative position of the gears; and the required amount of change serves as a measurement of the "ease-off." One set of bevel gears may be used as masters or "gages" to check other production gears. This running test quickly and easily assures quality and duplication for interchangeability, which is the ultimate goal of gear manufacture.

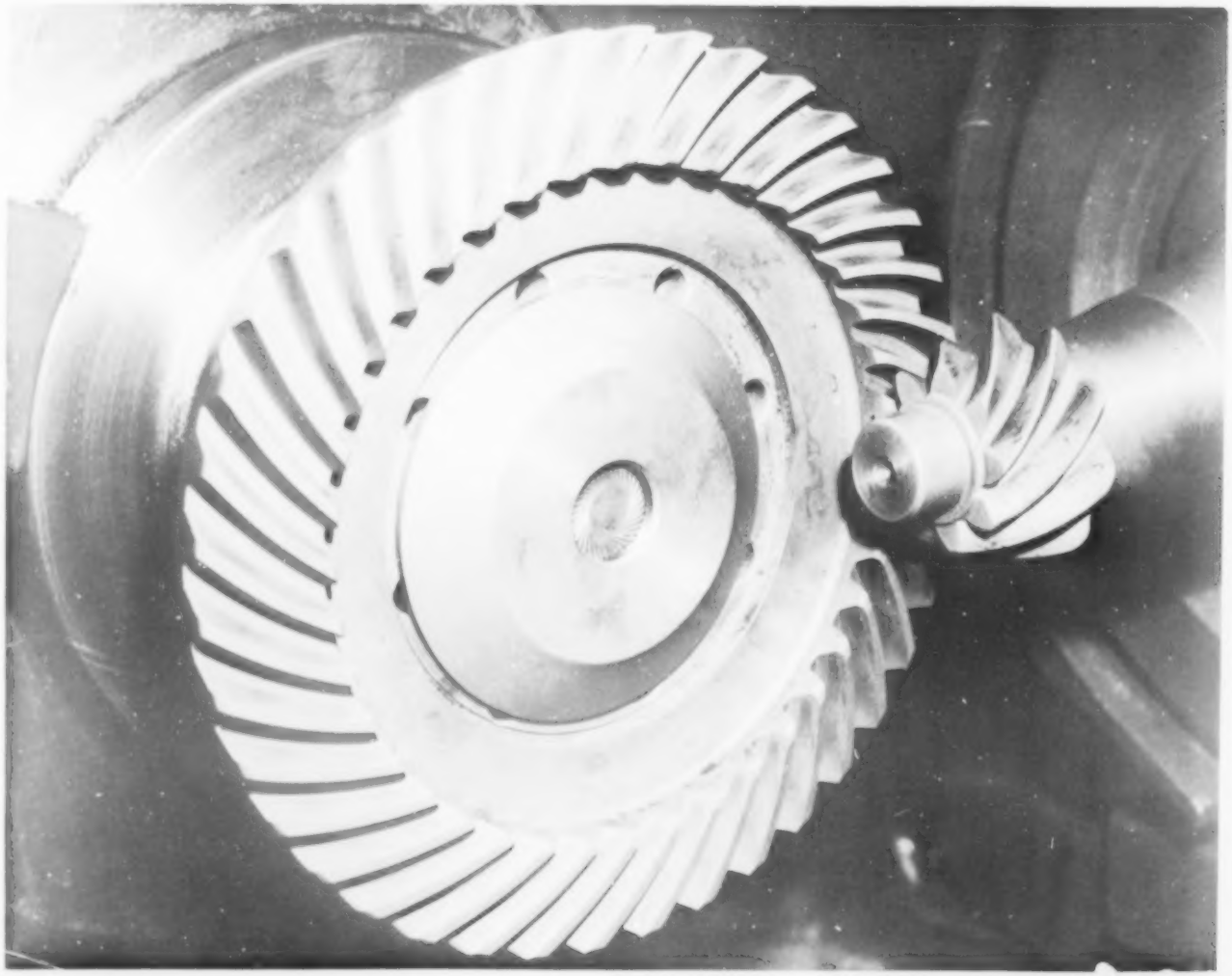


Figure 14 - Tooth Bearing in Spiral Bevel Gears

REFERENCES

1. Sang, Edward. "A New General Theory of the Teeth of Wheels," Edinburgh, 1852.

2. U.S. Patent 1,325,784 filed 1913, issued 1919.

3. Tredgold, Thomas. English civil engineer, 1788-1829. His approximation is named in most text books on gears.

4. Euler's theorem and Meunier's theorem. See Frost, "Solid Geometry," Third (and last) Edition. London: Macmillian, 1886, pp. 280, 282. Also James' and James' Mathematics Dictionary. Van Nuys, California: The Digest Press, 1942.